

**DESARROLLO DE UN MODELO MATEMÁTICO PARA DESCRIBIR EL
CRECIMIENTO DE ARÁNDANOS BILOXI MEDIANTE COMPARACIÓN DE
VARIABLES AMBIENTALES EN UN SISTEMA CERRADO Y ABIERTO**

AUTOR(ES)

JHONATAN MURCIA POSSO

ARIEL CAMILO SÁNCHEZ LÓPEZ

UNIVERSIDAD DE CUNDINAMARCA

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

FACATATIVÁ, 23 DE MAYO DEL 2025

**DESARROLLO DE UN MODELO MATEMÁTICO PARA DESCRIBIR EL
CRECIMIENTO DE ARÁNDANOS BILOXI MEDIANTE COMPARACIÓN DE
VARIABLES AMBIENTALES EN UN SISTEMA CERRADO Y ABIERTO**

AUTOR(ES)

DIRECTOR: ING. OSCAR JAVIER MORERA ZARATE

COINVESTIGADORES: ING. GINA MARIBEL VALENZUELA SABOGAL

JHONATAN MURCIA POSSO

ARIEL CAMILO SÁNCHEZ LÓPEZ

**GRUPO DE INVESTIGACIÓN DE SISTEMAS Y TECNOLOGÍA DE
FACATATIVÁ**

(GISTFA)

UNIVERSIDAD DE CUNDINAMARCA

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS

FACATATIVÁ, 23 DE MAYO DEL 2025

Nota de aceptación:

Firma del director de tesis

Firma del jurado

Firma del jurado

DEDICATORIA JHONATAN MURCIA POSSO

Dedico este proyecto principalmente a mis padres, hermanos, mi mascota y mi grupo de amigos y trabajo de la universidad quienes han sido un pilar fundamental para poder convertirme en la persona y profesional que hoy por hoy soy. Por ultimo y no menos importante también a mi director de proyecto y codirectora quienes ayudaron a el desarrollo de este proyecto de investigación

Por último, dedico el trabajo a todas aquellas personas que usan su conocimiento para poder crear formas o herramientas que hacen del mundo un lugar mejor habitable para todos los seres vivos, que tienen presente que antes de ser un gran profesional hay que ser un buen ser humano.

DEDICATORIA ARIEL CAMILO SÁNCHEZ LÓPEZ

A mis padres, hermanos y abuelos, presentes y ausentes, cuyo amor y apoyo han sido la fuerza que me ha impulsado en cada paso. A mis compañeros y amigos que conocí en la universidad, por las largas charlas y el buen humor. Y especialmente a mi fiel Sapita, que siempre me hace sonreír en los días de más esfuerzo.

AGRADECIMIENTOS JHONATAN MURCIA POSSO

Agradezco a mi director de tesis, coinvestigadora, jurados y al profesor Cesar Barahona quienes apoyaron a organizar y estructurar bien el proyecto en la parte documental y estructural, en conjunto a todos los docentes que construyeron el pensamiento crítico y la capacidad de poder razonar, quienes formaron experiencias enriquecedoras que me ayudan a crecer profesional y personalmente. A la Universidad de Cundinamarca por ser ese pilar de conocimiento que junto a sus aulas permitió enriquecer mi conocimiento.

De la misma manera agradezco a mis amigos y grupo de trabajo quienes me brindaron el apoyo para el desarrollo del proyecto, a mi amigo y compañero de tesis Ariel Sánchez por tener esa responsabilidad y sentido de compromiso con el proyecto, gracias a cada uno de ustedes porque sin ustedes no se hubiere podido concretar el desarrollo de este proyecto.

AGRADECIMIENTOS ARIEL CAMILO SÁNCHEZ LÓPEZ

Quiero agradecer a mi director, Óscar, y a mi coinvestigadora, Gina, por su apoyo constante al proyecto y sus valiosas aportaciones. Asimismo, agradezco a Gladys por su disposición a siempre ayudar. A los profesores quienes con su dedicación hicieron de cada clase un reto enriquecedor, y a la institución por brindar las condiciones necesarias para que este trabajo fuera posible.

COMPROMISO DE AUTOR

Yo, **Ariel Camilo Sánchez López** con célula de identidad No. **1007534897** y con cód. **461220260** estudiante del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, declaro que: El contenido del presente documento es un reflejo de mi trabajo personal y manifiesto que, ante cualquier notificación de plagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

Firma:

Camilo Sanchez

COMPROMISO DE AUTOR

Yo, **Jhonatan Murcia Posso** con célula de identidad No. **1006502314** y con cód. **461220228** estudiante del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, declaro que: El contenido del presente documento es un reflejo de mi trabajo personal y manifiesto que, ante cualquier notificación de plagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

Firma:

Jhonatan Murcia P.

Resumen

Esta investigación presenta el desarrollo de un modelo matemático orientado a describir la altura de la planta de arándanos, específicamente en su variedad Biloxi. El objetivo principal es brindar una herramienta que permita facilitar el control del cultivo mediante datos recopilados a través de un prototipo computacional, permitiendo identificar qué variables influyen de manera más significativa en su desarrollo, de la misma forma poder brindar bases para el desarrollo de herramientas más robustas que ya no solo permitan la descripción sino la predicción del crecimiento de la planta en general teniendo presentes las variables ambientales del entorno en donde se encuentren.

En la primera fase del proyecto, se exploraron distintos modelos matemáticos, destacándose principalmente tres: **los modelos aditivos generalizados (GAM)**, que utilizan splines cúbicos para capturar relaciones no lineales entre variables; **los modelos de series de tiempo**, que consideran el estado previo de la planta para comprender mejor su comportamiento en distintos ambientes; y **los modelos basados en curvas sigmoideas (curva S)**, frecuentemente asociados al patrón de crecimiento de las plantas.

Durante la segunda fase, se inició la recolección de datos mediante sensores conectados a un servidor MQTT, durante este proceso por limitaciones externas del proyecto se optó por la generación de datos sintéticos basados en una masa pequeña de datos reales para poder construir un volumen aceptable de información en datasets para el posterior entrenamiento del modelo. La implementación se realizó utilizando Python para el desarrollo del modelo matemático y JavaScript con React.js para la interfaz del software.

Finalmente, se presentan los resultados obtenidos en un módulo web de reportes, donde se detallan los modelos matemáticos utilizados, las variables estudiadas y los

resultados alcanzados. Asimismo, se incluyen los conjuntos de datos empleados para alimentar los modelos. Por último, se especifican los rangos adecuados en los que cada variable debe encontrarse para poder brindar un mayor aporte para la altura de la planta, respaldados por el análisis realizado con el modelo GAM.

Palabras clave: Modelo matemático, Crecimiento vegetal, Arándanos biloxi, Análisis de datos.

Abstract

This research presents the development of a mathematical model oriented to describe the height of the blueberry plant, specifically in its Biloxi variety. The main objective is to provide a tool to facilitate the control of the crop by means of data collected through a computational prototype, allowing to identify which variables have a significant influence on its development, and also to provide the basis for the development of more robust tools that not only allow the description but also the prediction of the growth of the plant in general, taking into account the environmental variables of the environment where they are located.

In the first phase of the project, different mathematical models were explored, mainly three: generalized additive models (GAM), which use cubic splines to capture nonlinear relationships between variables; time series models, which consider the previous state of the plant to better understand its behavior in different environments; and models based on sigmoid curves (S-curve), frequently associated with the growth pattern of plants.

During the second phase, data collection was initiated using sensors connected to an MQTT server; during this process, due to external limitations of the project, it was decided to generate synthetic data based on a small mass of real data in order to build an acceptable volume of information in datasets for the subsequent training of the model. The implementation was carried out using Python for the development of the mathematical model and JavaScript with React.js for the software interface.

Finally, the results obtained are presented in a web reporting module, where the mathematical models used, the variables studied, and the results achieved are detailed. The data sets used to feed the models are also included. Finally, the appropriate ranges in which each variable should be found in order to provide a greater contribution to plant height are specified, supported by the analysis carried out with the GAM model.

Key words: Mathematical model, Plant growth, Computational prototype, Biloxi blueberries, Data analysis.

Índice

Resumen.....	9
Abstract.....	10
Índice de figuras.....	17
Introducción	21
1. Informe de Investigación	23
1.1. Estado del Arte.....	23
1.2. Línea de Investigación	47
1.3. Planteamiento del Problema y Pregunta de Investigación	47
1.3.1. Planteamiento del Problema.....	47
1.3.2. Pregunta de Investigación	48
1.4. Objetivo general y objetivos específicos.....	49
1.4.1. Objetivo general.....	49
1.4.2. Objetivos específicos	49
1.5. Alcance e Impacto del Proyecto.....	49
1.6. Metodología.....	51
1.7. Marcos de Referencia	52
1.7.1. Marco teórico	52
1.7.2. Marco legal	64
2. Documentación del Software.....	73
2.1. Plan de Proyecto	73
2.2. Determinación de Requerimientos	77
Especificación del Diseño.....	85
2.2.1. Diagramas de Casos de Uso.....	85
2.2.2. Diagramas de Secuencia	90

2.2.3.	Diagramas de Actividades	104
2.3.	Diseños de Casos de Prueba	114
2.4.	Pruebas de Estrés y Carga (JMETER)	114
2.5.	Pruebas de Calidad y Rendimiento (SONARQUBE)	118
2.5.1.	Frontend	118
2.5.2.	Backend.....	125
2.6.	Estimación por Puntos de Función	126
2.7.	Integración de Datos	131
2.7.1.	Ambiente y Plantas	131
2.7.2.	Generación de Datos Sintéticos.....	132
2.7.3.	Modelo ARIMA y SARIMA.....	134
2.7.4.	Estructura Temporal.....	135
2.7.5.	Preprocesamiento Datos Reales	138
2.7.6.	Datasets para Modelo SARIMA	161
2.7.7.	Backcasting de Datos Sintéticos ARIMA	163
2.7.8.	Backcasting de Datos Sintéticos con red neuronal.....	165
2.7.9.	Variable Dependiente (altura de la planta).....	169
2.8.	Análisis de Modelos.....	173
2.8.1.	Modelo Aditivo GAM.....	173
2.8.2.	Modelo GAM para el Crecimiento de las Plantas.....	174
2.8.3.	Modelos no Lineales para Curva Sigmoide	175
2.8.4.	Beneficios del Modelo Sigmoide en la Descripción del Crecimiento Vegetal	176

	14
2.8.5. Análisis de Series de Tiempo	176
2.8.7. Redes Neuronales Recurrentes RNN	178
2.8.8. Funcionamiento de las Redes Neuronales Recurrentes.....	179
2.8.9. Aplicación Redes Neuronales Recurrentes en el Crecimiento de las Plantas	180
2.8.10. ¿Qué es Normalización de Datos?	180
2.8.11. Importancia de la Normalización en Modelos Predictivos del Crecimiento de Plantas	181
2.9. Desarrollo, Ajuste y análisis de los resultados obtenidos por los modelos ...	183
2.9.1. GAM	183
2.9.2. Análisis de Resultados GAM.....	186
2.9.3. Correlación de Kendall para Determinar la Relación entre las Variables Evaluada por GAM	201
2.9.4. Conclusiones de los Resultados Obtenidos del Modelo GAM	203
2.9.5. Modelos de Curva Sigmoide (S).....	205
2.9.6. Función Gompertz.....	205
2.9.7. Función Logística.....	209
2.9.8. Comparativa entre Gompertz y Logística	213
2.10. Resultados.....	214
Resultados: Integración de datos.....	214
Resultados: Desarrollo modelo matemático.....	216
2.11. Conclusiones y Recomendaciones	223

	15
Conclusiones	223
Recomendaciones.....	225
3. Bibliografía	227
4. Anexos	245

Índice de Tablas

Tabla 1 Artículos de uso GAM.....	30
Tabla 2 Ejemplos de uso de MANOVA y ANOVA	41
Tabla 3 Descripción de las tareas del proyecto	75
Tabla 4 Definición y Acrónimos	78
Tabla 5 Regencias Estándar de los Requerimientos	79
Tabla 6 Descripción Rol Super Administrador.....	80
Tabla 7 Descripción Rol Usuario	80
Tabla 8 Requerimiento RF-01.....	82
Tabla 9 Requerimiento RF-02.....	82
Tabla 10 Requerimiento RF-03.....	83
Tabla 11 Requerimiento RF-04.....	83
Tabla 12 Requerimiento RF-05.....	83
Tabla 13 Tabla descriptiva de los casos de uso para el rol de Super Administrador.	86
Tabla 14 Tabla descriptiva de los casos de uso para el rol de Usuario.....	89
Tabla 15 Descripción de los componentes de un diagrama de secuencia	90
Tabla 16 Descripción de los componentes	104

Tabla 17 Pruebas SonarQube	118
Tabla 18 reportService.Js.....	120
Tabla 19 Pesos de asignación según international function pont users Group (IFPUG)	126
Tabla 20 Identificación y clasificación de unidades funcionales	127
Tabla 21 Cálculo de grado de influencia	128
Tabla 22 Fecha y minutos transcurridos desde la siembra	137
Tabla 23 Estructura dataset para entrenar modelo de datos sintéticos	162
Tabla 24 Estructura dataset para predecir modelo de datos sintéticos	163
Tabla 25 Registros de altura de las plantas en centímetros.	169
Tabla 26 Estructura dataset para entrenamiento de modelos.....	172
Tabla 27 Tabla de características de modelos.....	182
Tabla 28 Descripciones de variables predictoras y categóricas.....	185
Tabla 29 Resultado de métricas.....	186
Tabla 30 Resultado de la regresión con LinearGAM (Generalized Additive Model)	187
Tabla 31 Aporte de las variables al modelo.....	187
Tabla 32 Resultados de ajuste del modelo Gompertz.....	208
Tabla 33 Resultados de ajuste del modelo logístico.....	211
Tabla 34 Rangos óptimos de las Variables	224

Índice de figuras

Figura 1 Agricultura de Precisión	53
Figura 2 Agricultura tradicional	57
Figura 3 Agricultura industrial.	58
Figura 4 Agricultura Vertical	60
Figura 5 Imagen de tareas con sus respectivas sucesoras en proyect libre	73
Figura 6 Diagrama Gantt en formato de red.	74
Figura 7 Diagrama de casos de uso para el rol de Super Usuario.....	85
Figura 8 Diagrama de casos de uso para el rol de Usuario.	88
Figura 9 Diagrama de secuencia Generar reportes	92
Figura 10 Diagrama de secuencia para ver interfaz.....	93
Figura 11 Diagrama de secuencia para visualización dataset.	94
Figura 12 Diagrama de secuencia: Técnicas Metodologías y técnicas usadas para la elaboración del modelo.	95
Figura 13 Diagrama de secuencia para visualización modelo matemático.....	96
Figura 14 Diagrama de secuencia para Crear Dataset	97
Figura 15 Diagrama de secuencia para Leer Dataset.....	99
Figura 16 Diagrama de secuencia para Actualizar Dataset.....	101
Figura 17 Diagrama de secuencia Eliminar Dataset	103
Figura 18 Diagrama de actividad para generar reportes.	105
Figura 19 Diagrama de actividad para ver interfaz web.	106
Figura 20 Diagrama de actividad para visualizar dataset	107
Figura 21 Diagrama de actividad para la visualización metodologías y técnicas.	108
Figura 22 Diagrama de actividad para la visualización del modelo matemático... ..	109

Figura 23 Diagrama de actividad para crear el dataset	110
Figura 24 Diagrama de actividad para leer el dataset	111
Figura 25 Diagrama de actividad para Actualizar dataset.....	112
Figura 26 Diagrama de actividad para eliminar el dataset.....	113
Figura 27 Configuración Thread Group (20 hilos 4 segundos)	115
Figura 28 Resultados Obtenidos	116
Figura 29 Configuración Thread Group (100 hilos 20 segundos)	117
Figura 30 Resultados Obtenidos	117
Figura 31 Resultado General de calidad	121
Figura 32 Duplicado de Código.....	122
Figura 33 Flexibilidad / Confiabilidad.....	123
Figura 34 Mantenibilidad.....	124
Figura 35 Seguridad.....	125
Figura 36 Estructura archivo CSV con fechas y minutos desde la siembra inicial.	138
Figura 37 Grafica de porcentaje de humedad relativa vs tiempo del grupo de sensores 1 ambiente semi controlado.	141
Figura 38 Porcentaje humedad del sustrato vs tiempo del grupo de sensores 1 ambiente semi controlado.....	142
Figura 39 Datos ADC de Luminosidad vs tiempo real para el grupo de sensores 1 ambiente semi controlado.....	143
Figura 40 Datos de grados temperatura real vs tiempo para el grupo de sensores 1 ambiente semi controlado.....	144

Figura 41 Datos de rayos UV vs el tiempo para el grupo de sensores 1 ambiente semi controlado.	145
Figura 42 Datos del porcentaje de Humedad relativa vs el tiempo para el grupo de sensores 2 ambiente semi controlado.	146
Figura 43 Datos del porcentaje de Humedad del sustrato vs el tiempo grupo de sensores 2 ambiente semi controlado.	147
Figura 44 Datos ADC de Luminosidad vs el tiempo para el grupo de sensores 2 ambiente semi controlado.	148
Figura 45 Datos de grados de temperatura vs el tiempo para el grupo de sensores 2 ambiente semi controlado.	149
Figura 46 Datos de rayos UV vs el tiempo real para el grupo de sensores 2 ambiente semi controlado.	150
Figura 47 Diagrama de caja sobre los datos de porcentaje de humedad relativa real para el grupo de sensores 1 ambiente semi controlado.	151
Figura 48 Histograma porcentaje de humedad relativa grupo de sensores 1 ambiente semi controlado.	152
Figura 49 Diagrama de caja porcentaje de humedad de sustrato grupo de sensores 1 ambiente semi controlado.	153
Figura 50 Histograma humedad de sustrato grupo de sensores 1 ambiente semi controlado.	154
Figura 51 Temperatura vs Humedad de Sustrato grupo de sensores 1 ambiente semi controlado.	155
Figura 52 Diagrama de caja ADC de luminosidad real grupo de sensores 1 ambiente semi controlado.	156

Figura 53 Histograma luminosidad real grupo de sensores 1 ambiente semi controlado.	157
Figura 54 Diagrama de caja temperatura real grupo de sensores 1 ambiente semi controlado.	158
Figura 55 Histograma temperatura real grupo de sensores 1 ambiente semi controlado.	159
Figura 56 Diagrama de caja UV real grupo de sensores 1 ambiente semi controlado.	160
Figura 57 Histograma UV real grupo de sensores 1 ambiente semi controlado.	160
Figura 58 Predicciones SARIMAX con dos días de datos de entrenamiento.	164
Figura 59 Predicciones ARIMA para dos días anteriores.	165
Figura 60 Datos sintéticos y reales de grados de temperatura	166
Figura 61 Datos sintéticos y reales de porcentaje de humedad relativa.	167
Figura 62 Datos sintéticos y reales de porcentaje de humedad de sustrato.	168
Figura 63 Datos sintéticos y reales de luminosidad.	169
Figura 64 Altura de las plantas.	171
Figura 65 Grafica de efecto parcial de temp_high S(0) vs altura	189
Figura 66 Grafica de efecto parcial de temp_low S(1)	191
Figura 67 Grafica de efecto parcial de hum_rel S(2).....	192
Figura 68 Gráfica de efecto parcial de min_siembra S(3)	193
Figura 69 Gráfica de efecto parcial de hum_sustr S(4).....	194
Figura 70 Gráfica de efecto parcial de lum S(5).....	196
Figura 71 Grafica de efecto parcial de temp S(6)	197
Figura 72 Gráfica poda_engrosamiento (Variable categórica poda f(7)).....	198

Figura 73 Gráfica de altura predicha vs altura real.....	199
Figura 74 Gráfica de tiempo vs altura predicha y altura real.....	200
Figura 75 Matriz de correlación Kendall.....	202
Figura 76 Gráfica de distribución de los errores (Residuos).....	204
Figura 77 Ajuste de función Gompertz a datos de altura con offset.....	206
Figura 78 Ajuste de función Gompertz a datos de altura con offset y datos enriquecidos.....	208
Figura 79 Ajuste de función logística a datos de altura con offset y datos enriquecidos.....	211
Figura 80 Gráfica de comparativa entre Gompertz y Logística.....	213
Figura 81 Información sobre el dataset obtenido.....	215
Figura 82 Resultado Modulo web.....	221

Introducción

En el presente libro de grado se describió la investigación realizada y los resultados obtenidos durante el desarrollo del proyecto, cuyo objetivo fue describir la altura de las plantas mediante modelos matemáticos. Además, de las bases para el desarrollo de herramientas más robustas para la toma de decisiones basada en datos.

Inicialmente, se llevó a cabo una revisión teórica donde se presentaron los principales modelos matemáticos considerados, así como técnicas estadísticas complementarias y las variables ambientales evaluadas. Posteriormente, se abordó el

desarrollo del software junto con las pruebas correspondientes, incluyendo una planificación detallada del proyecto mediante herramientas de gestión del tiempo.

Como fase principal se optó por el desarrollo de métodos que permitieran generar datos sintéticos basados en masas pequeñas de datos, ya que no se contaba con un nivel de información suficiente para poder alimentar a los modelos, se evaluaron diferentes modelos como ARIMA y SARIMAX los cuales no dieron resultados adecuados lo que llevó a emplear una red neuronal que permitiera capturar los patrones de cada una de las variables estudiadas. Por medio de los resultados obtenidos se construyó el dataset con el cual se entrenó los modelos.

Los modelos obtenidos permitieron analizar los rangos en donde cada una de estas variables tenía un mayor aporte, facilitando la comprensión del efecto que esta tenía respecto a la variable objetivo en este caso la altura. El proyecto se llevó a cabo en un ambiente semicontrolado con dos grupos de sensores y 3 plantas.

Se seleccionaron dos enfoques matemáticos principales para describir el crecimiento de la planta: un modelo GAM (Generalized Additive Model), el cual permitió estimar y visualizar el efecto parcial e independiente de cada variable ambiental sobre la altura de la planta; y dos modelos de tipo sigmoide (Gompertz y logístico), utilizados para determinar parámetros clave como la tasa de crecimiento y punto de curvatura. Estos resultados fueron acompañados de representaciones gráficas para facilitar su interpretación.

Finalmente, se realizó un análisis mediante la correlación de Kendall con el fin de identificar la posible colinealidad entre variables. Los resultados indicaron que cada variable aportó información relevante e independiente en relación con la altura de las

plantas. El libro concluye con los resultados obtenidos, las conclusiones derivadas y recomendaciones prácticas surgidas del proceso investigativo, complementado con anexos técnicos adicionales.

1. Informe de Investigación

1.1. Estado del Arte

Variables que afectan el crecimiento de las plantas

Luz:

La luz es un factor crítico en el crecimiento de las plantas, afectando múltiples aspectos de su desarrollo.

- **Intensidad de la luz:** Según Poorter et al. (2019), la intensidad de la luz influye directamente en la tasa fotosintética. Las plantas expuestas a alta intensidad lumínica tienden a desarrollar hojas más gruesas y compactas, con mayor densidad de cloroplastos, lo que aumenta su capacidad fotosintética. En contraste, las plantas en condiciones de baja luz suelen tener hojas más grandes y delgadas para maximizar la captación de luz.
- **Duración de la luz (fotoperiodo):** El fotoperiodo regula procesos como la floración y la dormancia. Song et al. (2015) explican que muchas plantas han evolucionado para florecer en respuesta a cambios en la duración del día, un proceso conocido como fotoperiodismo. Por ejemplo, las plantas de día corto florecen cuando la duración de la noche supera un umbral crítico, mientras que las plantas de día largo requieren un período de luz más prolongado.

- **Calidad de la luz:** Diferentes longitudes de onda afectan distintos aspectos del crecimiento. Ouzounis et al. (2015) detallan que la luz azul promueve la fotomorfogénesis, incluyendo la inhibición de la elongación del tallo, la expansión de las hojas y la producción de clorofila. La luz roja y roja lejana influyen en la germinación de semillas, la elongación del tallo y la floración a través del sistema de fitocromos.

Temperatura:

La temperatura afecta casi todos los procesos metabólicos en las plantas.

- **Crecimiento vegetativo:** Hatfield y Prueger (2015) indican que cada especie tiene un rango de temperatura óptimo para el crecimiento. Fuera de este rango, el crecimiento se ralentiza. Las temperaturas muy altas pueden causar daños en las proteínas y membranas celulares, mientras que las temperaturas muy bajas pueden provocar la formación de cristales de hielo intracelulares, dañando los tejidos.
- **Fotosíntesis y respiración:** Según Way y Yamori (2014), la tasa fotosintética aumenta con la temperatura hasta un punto óptimo, después del cual comienza a disminuir. La respiración, por otro lado, continúa aumentando con la temperatura, lo que puede resultar en un balance negativo de carbono a temperaturas muy altas.
- **Desarrollo reproductivo:** Prasad et al. (2017) señalan que las altas temperaturas durante la floración pueden reducir la viabilidad del polen y la receptividad del estigma, afectando la formación de semillas y frutos. Las bajas temperaturas pueden retrasar la floración y el desarrollo de frutos.

Agua:

El agua es esencial para numerosos procesos fisiológicos en las plantas.

- **Transporte de nutrientes:** El agua es el medio de transporte para los nutrientes en la planta. Según Fahad et al. (2017), el estrés hídrico limita la absorción y el transporte de nutrientes, lo que puede resultar en deficiencias nutricionales incluso cuando los nutrientes están presentes en el suelo.
- **Fotosíntesis y transpiración:** El estrés hídrico provoca el cierre de las estomas para conservar agua, lo que también reduce la entrada de CO₂, disminuyendo la tasa fotosintética. Flexas et al. (2018) detallan cómo el estrés hídrico afecta los componentes bioquímicos y difusionales de la fotosíntesis.

Nutrientes del suelo:

Los nutrientes son componentes esenciales para el crecimiento y desarrollo de las plantas.

- **Macronutrientes:** White et al. (2012) describen cómo los macronutrientes (N, P, K, Ca, Mg, S) son necesarios en grandes cantidades. Por ejemplo, el nitrógeno es crucial para la formación de clorofila y proteínas, el fósforo es esencial para la transferencia de energía y la formación de ácidos nucleicos, y el potasio es importante para la regulación osmótica y la activación enzimática.
- **Micronutrientes:** Aunque requeridos en cantidades menores, los micronutrientes (Fe, Mn, Zn, Cu, B, Mo, Cl) son igualmente importantes. Hänsch y Mendel (2009) explican que estos elementos son a menudo cofactores de enzimas críticas. Por ejemplo, el hierro es esencial para la síntesis de clorofila y el funcionamiento de los fotosistemas.

- **Interacciones entre nutrientes:** La disponibilidad y absorción de un nutriente puede afectar la de otros. Rietra et al. (2017) discuten cómo, por ejemplo, altos niveles de fósforo pueden inhibir la absorción de zinc y hierro.

CO₂:

El dióxido de carbono es el sustrato principal para la fotosíntesis.

- **Efecto fertilizante del CO₂:** Ainsworth y Long (2005) reportan que el aumento de CO₂ atmosférico generalmente incrementa la tasa fotosintética y el crecimiento en plantas C₃. Sin embargo, este efecto puede ser limitado por otros factores como la disponibilidad de nutrientes o agua.
- **Aclimatación:** Leakey et al. (2009) explican que la exposición prolongada a elevado CO₂ puede llevar a una aclimatación de la fotosíntesis, donde la planta reduce su capacidad fotosintética para equilibrar su metabolismo de carbono y nitrógeno.

Humedad:

La humedad del aire afecta la tasa de transpiración y, por ende, la absorción de agua y nutrientes.

- **Déficit de presión de vapor (VPD):** Will et al. (2013) explican que un alto VPD (baja humedad relativa) aumenta la tasa de transpiración, lo que puede llevar a estrés hídrico si no hay suficiente agua disponible en el suelo. Esto puede resultar en el cierre de estomas, reduciendo la fotosíntesis y el crecimiento.
- **Enfermedades fúngicas:** Según Juroszek y von Tiedemann (2015), la alta humedad puede promover el desarrollo de enfermedades fúngicas, afectando negativamente el crecimiento y la salud de las plantas.

- **Polinización:** La humedad también afecta la polinización. Ceccon et al. (2019) indican que una humedad muy alta o baja puede reducir la viabilidad del polen y la receptividad del estigma en algunas especies.

pH del suelo:

El pH del suelo influye en la disponibilidad de nutrientes y la actividad microbiana.

- **Disponibilidad de nutrientes:** Gentili et al. (2018) explican que diferentes nutrientes tienen distintas disponibilidades según el pH del suelo. Por ejemplo, el hierro, manganeso, cobre y zinc son más disponibles en suelos ácidos, mientras que el molibdeno es más disponible en suelos alcalinos.
- **Toxicidad por aluminio:** Según Kochian et al. (2015), en suelos muy ácidos (pH < 5.5), el aluminio puede volverse soluble y tóxico para las raíces, inhibiendo su crecimiento y la absorción de agua y nutrientes.
- **Actividad microbiana:** Rousk et al. (2009) indican que el pH del suelo afecta la composición y actividad de las comunidades microbianas del suelo, lo que a su vez influye en la descomposición de la materia orgánica y el ciclo de nutrientes.

Modelos estadísticos que ayudan a describir el crecimiento de plantas por medio de variables ambientales

Modelo de von Bertalanffy:

Originalmente desarrollado para describir el crecimiento animal, pero también aplicable a plantas.

$$Y = Y_{\infty} * (1 - \exp(-k * (t - t_0)))^3$$

Donde Y es la variable de crecimiento, Y_{∞} es el tamaño máximo teórico, k es la tasa de crecimiento, t es el tiempo, y t_0 es el tiempo teórico cuando $Y = 0$.

Shi et al. (2013) aplican este modelo al crecimiento de árboles en bosques.

Shi, P., Körner, C., & Hoch, G. (2013). A test of the growth-limitation theory for alpine tree line formation in evergreen and deciduous taxa of the eastern Himalayas. *Functional Ecology*, 20(2), 213-220. <https://edoc.unibas.ch/8479/>

Modelo polinomial:

Los modelos polinomiales, especialmente los cuadráticos y cúbicos, pueden ser útiles para describir el crecimiento de plantas en períodos cortos o cuando la forma de la curva de crecimiento es compleja.

$$Y = a + bt + ct^2$$

Donde Y es la variable de crecimiento, a , b , y c son constantes, y t es el tiempo.

Archontoulis y Miguez (2015) Ellos nos presentan la parte de como mucho de los procesos relacionados con los cultivos están mejor descritos por modelos no lineales que por los lineales, donde se menciona también el apartado que dependiendo el objetivo que se tenga van a variar los parámetros, las prioridades, en donde la mayoría se basa en el tiempo, aclarando que un buen ajuste del modelo dará un buen resultado cercano al esperado.

Según Archontoulis y Miguez (2015) “Proponemos pasos para ajustar modelos no lineales como se describe en un diagrama de flujo y analizamos cada paso por separado proporcionando ejemplos y actualizaciones sobre los procedimientos utilizados. Se

consideran los siguientes pasos: (i) elegir modelos candidatos, (ii) establecer valores iniciales, (iii) ajustar modelos, (iv) verificar la convergencia y las estimaciones de parámetros, (v) encontrar el "mejor" modelo entre los modelos competidores, (vi) comprobar los supuestos del modelo (análisis residual) y (vii) calcular los descriptores estadísticos y los intervalos de confianza. También se abordan los mecanismos de retroalimentación asociados (es decir, la homogeneidad de la varianza del modelo).”, en donde se le da mayor importancia a el primer paso que es seleccionar modelos candidatos, y en el artículo se proponen 77 ecuaciones diferentes con el significado a los parámetros asociados además de esto ejemplos de aplicaciones en agricultura de cada uno de estos.

Modelos GAM para la descripción del crecimiento de las plantas

Cómo los GAM ayudan a describir el crecimiento de las plantas:

Flexibilidad en la modelación de relaciones no lineales: Los GAM permiten modelar relaciones complejas y no lineales entre variables predictoras y el crecimiento de las plantas sin asumir una forma funcional específica (Wood, 2017). Esto es particularmente útil en estudios ecológicos donde las relaciones entre variables ambientales y el crecimiento de las plantas son a menudo no lineales y complejas (Zuur et al., 2009).

Manejo de múltiples predictores: Los GAM pueden incorporar simultáneamente múltiples variables predictoras, permitiendo un análisis integral de los factores que afectan el crecimiento de las plantas (Pedersen et al., 2019). Esto es crucial en estudios ecológicos donde múltiples factores ambientales influyen en el crecimiento de las plantas.

Captura de efectos estacionales y temporales: Los GAM son eficaces para modelar patrones cíclicos o estacionales en el crecimiento de las plantas, que son comunes debido a los cambios anuales en temperatura, luz y otros factores ambientales (Wood, 2017). Esto permite a los investigadores capturar y analizar la dinámica temporal del crecimiento de las plantas.

Manejo de datos espaciales: Los GAM pueden incorporar componentes espaciales, lo que es útil para modelar el crecimiento de plantas en diferentes ubicaciones geográficas (Zuur et al., 2009). Esto permite a los investigadores analizar cómo varía el crecimiento de las plantas en diferentes escalas espaciales.

Visualización de relaciones: Los GAM proporcionan herramientas de visualización que permiten una interpretación intuitiva de las relaciones entre variables (Pedersen et al., 2019). Esto facilita la comunicación de resultados complejos y la identificación de patrones importantes en el crecimiento de las plantas.

Tabla 1
Artículos de uso GAM.

Título	Resumen	Detalles
Functional traits explain light and size response of growth rates in tropical tree species	Las relaciones entre los rasgos funcionales y las tasas demográficas promedio o potenciales han proporcionado información sobre las limitaciones funcionales y las compensaciones subyacentes a las estrategias de historia de vida de las especies de árboles tropicales. Hemos ampliado este marco de trabajo descomponiendo las tasas de crecimiento de ~130 000 árboles de 171 especies de árboles	Autores: Vásquez-Martínez, Lizarazo-Peña, Darghan, Nadja Rüger, Christian Wirth, S. Joseph Wright, Richard Condit

	<p>neotropicales en crecimiento intrínseco y la respuesta del crecimiento a la luz y el tamaño. Relacionamos estas características de crecimiento con múltiples rasgos funcionales (densidad de la madera, estatura adulta, masa de semillas, rasgos de las hojas) en un modelo bayesiano jerárquico que tuvo en cuenta el error de medición y la variabilidad intraespecífica de los rasgos funcionales. La densidad de la madera fue el rasgo más importante que determinó las tres características de crecimiento. Las tasas de crecimiento intrínsecas también estuvieron fuertemente relacionadas con la estatura adulta, mientras que todos los rasgos contribuyeron a la respuesta a la luz. Nuestro análisis produjo un modelo predictivo que permite la estimación de las características de crecimiento de especies raras sobre la base de unos pocos rasgos morfológicos fácilmente medibles.</p>	<p>Lugar: Ecological society of america</p> <p>Fecha de publicación: 2021</p>
<p>R package for studying linear and nonlinear responses between tree-rings and daily environmental data</p>	<p>En este artículo, presentamos el paquete R dendroTools para estudiar las relaciones estadísticas entre los parámetros de los anillos de los árboles y los datos ambientales diarios. La función principal del paquete es <code>daily_response()</code>, que funciona deslizando una ventana móvil a través de los datos ambientales diarios y calculando métricas estadísticas con uno o más indicadores de los anillos de los</p>	<p>Autores: Jernej Jevšenak, Tom Levanič</p> <p>Lugar: USA.</p> <p>Fecha de publicación: 2018</p>

árboles. Las métricas posibles son el coeficiente de correlación, el coeficiente de determinación y el coeficiente de determinación ajustado. Además de la regresión lineal, es posible utilizar una red neuronal artificial no lineal con el algoritmo de entrenamiento de regularización bayesiana (brnn). dendroTools brinda la oportunidad de utilizar datos climáticos diarios y funciones no lineales robustas para el análisis de las relaciones entre el clima y el crecimiento. De este modo, los modelos deberían adaptarse mejor al crecimiento real (continuo) de los árboles y deberían ganar en capacidad predictiva. El paquete R dendroTools está disponible de forma gratuita en el repositorio CRAN. La funcionalidad del paquete se demuestra en dos ejemplos, uno que utiliza una cronología de área media de los vasos (MVA) y otro un ancho de los anillos de los árboles tradicional (TRW).

Contrasting trait syndromes in angiosperms and conifers are associated with different	Estudios recientes a gran escala sobre el crecimiento de los árboles en la Península Ibérica han demostrado efectos positivos y negativos contrastantes de la temperatura en las angiospermas y coníferas mediterráneas. En este artículo, revisamos las diferentes hipótesis que pueden explicar estas tendencias y	Autores: Carnicer, J., Barbeta, A., Sperlich, D., Coll, M., & Peñuelas, J. (2013) Lugar: S.L.
--	--	--

responses of tree growth to temperature on a large scale	proponemos que las respuestas contrastantes observadas del crecimiento de los árboles a la temperatura en esta región podrían estar asociadas con un continuo de diferencias de caracteres entre angiospermas y coníferas. Las angiospermas y las coníferas difieren en los efectos de la fenología en su productividad, en su alometría de crecimiento y en su sensibilidad a la competencia. Además, las angiospermas y las coníferas difieren significativamente en los márgenes de seguridad hidráulica, la sensibilidad de la conductancia estomática al déficit de presión de vapor (VPD), la capacidad de recuperación del xilema o la tasa de transferencia de carbono. Estas diferencias podrían explicarse por características clave del xilema, como el contenido de carbohidratos no estructurales (NSC), la fracción de parénquima leñoso o la capacitancia de la madera. Sugerimos que las diferencias de caracteres revisadas definen dos estrategias ecofisiológicas contrastantes que pueden determinar respuestas de crecimiento cualitativamente diferentes al aumento de la temperatura y la sequía. La mejora de los experimentos de jardines comunes recíprocos a lo largo de gradientes altitudinales o latitudinales sería clave para cuantificar la importancia relativa de las diferentes hipótesis revisadas. Finalmente, mostramos que los impactos del calentamiento en esta área ocurren	Fecha de publicación: 2013
---	--	-----------------------------------

en un contexto ecológico caracterizado por el avance de la sucesión forestal y el aumento del predominio de los árboles angiospermas en áreas extensas. En este contexto, examinamos las relaciones empíricas entre las respuestas del crecimiento de los árboles a la temperatura y los márgenes de seguridad hidráulica en árboles angiospermas y coníferas. Nuestros hallazgos sugieren un escenario futuro en los bosques mediterráneos caracterizado por respuestas demográficas contrastantes en árboles coníferos y angiospermas tanto a la temperatura como a la sucesión forestal, con un mayor predominio de los árboles angiospermas e impactos particularmente negativos en los pinos.

Nota. En esta tabla se muestran artículos en los cuales se emplearon los modelos GAM para poder describir o predecir crecimiento de las plantas mencionadas en los estudios

Modelos Comparativos (Análisis de varianza ANOVA, Análisis discriminante, Análisis de varianza multivariable MANOVA)

Análisis de Varianza Anova.

El Análisis de Varianza (ANOVA) es una técnica estadística que se utiliza para comparar las medias de tres o más grupos y determinar si existen diferencias significativas entre ellos (Howell, 2012). ANOVA funciona dividiendo la variabilidad total observada en los datos en dos componentes: la variabilidad entre grupos y la variabilidad dentro de los grupos.

Si la variabilidad entre grupos es significativamente mayor que la variabilidad dentro de los grupos se concluye que hay diferencias significativas entre las medias de los grupos.

Si la variabilidad entre grupos no es significativamente mayor que la variabilidad dentro de los grupos se concluye que no hay evidencia suficiente para afirmar que las medias de los grupos son diferentes.

En un estudio hipotético, un investigador podría usar ANOVA para comparar el crecimiento en altura de una especie de planta bajo tres niveles diferentes de fertilizante (bajo, medio, alto). ANOVA permitiría determinar si las diferencias observadas en la altura media de las plantas entre los tres grupos son estadísticamente significativas.

O como se puede evidenciar en los siguientes estudios en donde se ha implementado ANOVA para el análisis de algunas plantas.

- a. Comparar tasas de crecimiento:** ANOVA puede determinar si existen diferencias significativas en las tasas de crecimiento de plantas sometidas a diferentes condiciones. Por ejemplo, Gianoli y González-Teuber (2005) utilizaron ANOVA para comparar las tasas de crecimiento de plantas de *Convolvulus chilensis* bajo diferentes niveles de luz y disponibilidad de agua.
- b. Evaluar el efecto de factores ambientales:** Se puede usar ANOVA para analizar cómo diferentes factores ambientales afectan el crecimiento de las plantas. Poorter y Nagel (2000) emplearon ANOVA en su meta-análisis para examinar cómo la disponibilidad de luz y nutrientes afecta la distribución de biomasa en plantas.

- c. Analizar respuestas a tratamientos:** ANOVA es útil para comparar las respuestas de crecimiento a varios tratamientos. Por ejemplo, Lambers et al. (2008) describen en su libro cómo ANOVA se usa para evaluar los efectos de diferentes niveles de nutrientes en el crecimiento de las plantas.
- d. Estudiar interacciones entre factores:** El ANOVA factorial permite a los investigadores examinar cómo múltiples factores interactúan para afectar el crecimiento de las plantas. Farooq et al. (2009) utilizaron ANOVA factorial para estudiar los efectos interactivos del estrés por sequía y la etapa de crecimiento en el rendimiento del arroz.
- e. Analizar componentes del crecimiento:** ANOVA puede aplicarse a varios componentes del crecimiento de las plantas, como altura, área foliar, biomasa, o tasa de fotosíntesis. Reich et al. (2003) emplearon ANOVA para analizar cómo la biomasa y otros rasgos de las plantas varían entre diferentes grupos funcionales y biomas.

Análisis Discriminante.

El análisis discriminante es una técnica estadística multivariante utilizada para clasificar objetos o individuos en grupos mutuamente excluyentes basándose en un conjunto de variables independientes (Härdle & Simar, 2015). Este método también se emplea para identificar las variables que mejor diferencian entre estos grupos y crear funciones predictivas para clasificar nuevos casos.

Según Tabachnick y Fidell (2019), el proceso implica:

- f.** Calcular las medias de grupo y las matrices de covarianza.

- g. Determinar qué variables discriminan entre los grupos.
- h. Establecer procedimientos para clasificar nuevos casos.

El análisis discriminante puede ayudar con la descripción y análisis del crecimiento de una planta de la siguiente manera:

- **Clasificación de etapas de crecimiento:** El análisis discriminante puede utilizarse para clasificar plantas en diferentes etapas de crecimiento basándose en múltiples variables morfológicas y fisiológicas (Gratani et al., 2014).
- **Identificación de factores críticos de crecimiento:** Puede ayudar a identificar qué variables (como altura, área foliar, biomasa) son más importantes para discriminar entre plantas con diferentes tasas de crecimiento o respuestas a tratamientos (Carvalho et al., 2019).
- **Predicción de respuestas a condiciones ambientales:** El análisis discriminante puede predecir cómo las plantas responderán a diferentes condiciones ambientales basándose en sus características de crecimiento (Pérez-Harguindeguy et al., 2013).
- **Diferenciación entre especies o variedades:** Puede utilizarse para distinguir entre diferentes especies o variedades de plantas basándose en sus patrones de crecimiento y características morfológicas (Klingenberg, 2016).
- **Evaluación de efectos de tratamientos:** El análisis discriminante puede ayudar a evaluar cómo diferentes tratamientos (por ejemplo, fertilizantes, riego) afectan el crecimiento de las plantas y clasificar las plantas según su respuesta a estos tratamientos (Poorter et al., 2012).

Ejemplo de aplicación

Gratani et al. (2014) utilizaron el análisis discriminante para estudiar las respuestas de crecimiento de dos especies de roble (*Quercus ilex* y *Q. pubescens*) a diferentes condiciones de luz. Las variables utilizadas incluyeron área foliar específica, contenido de clorofila, y tasa de asimilación neta.

Análisis de varianza multivariable MANOVA.

El Análisis de Varianza Multivariable (MANOVA, por sus siglas en inglés) es una técnica estadística que extiende el análisis de varianza (ANOVA) a situaciones donde hay múltiples variables dependientes que están relacionadas entre sí. Voy a proporcionarte una explicación detallada del MANOVA, junto con sus aplicaciones y consideraciones importantes. El MANOVA es una técnica estadística utilizada para determinar si existen diferencias significativas entre grupos en múltiples variables dependientes simultáneamente. Es una extensión del ANOVA que permite analizar el efecto de una o más variables independientes (factores) sobre dos o más variables dependientes correlacionadas (Tabachnick & Fidell, 2013).

Características principales del MANOVA:

- Múltiples variables dependientes: A diferencia del ANOVA, que analiza una sola variable dependiente, el MANOVA puede manejar múltiples variables dependientes simultáneamente.
- Control del error Tipo I: Al analizar múltiples variables dependientes en una sola prueba, el MANOVA ayuda a controlar la tasa de error Tipo I que podría inflarse si se realizaran múltiples ANOVA separados (Huberty & Morris, 1989).

- Consideración de las correlaciones: El MANOVA tiene en cuenta las correlaciones entre las variables dependientes, lo que puede revelar efectos que no serían detectados por ANOVA separados.

Consideraciones importantes:

- **Tamaño de la muestra:** El MANOVA requiere un tamaño de muestra relativamente grande, especialmente cuando se aumenta el número de variables dependientes (Field, 2013).
- **Normalidad multivariada:** El MANOVA asume que las variables dependientes siguen una distribución normal multivariada (Tabachnick & Fidell, 2013)

El MANOVA puede ayudar de las siguientes formas en la descripción del crecimiento de plantas con base a sus variables ambientales de la siguiente forma

- **Análisis simultáneo de múltiples variables de crecimiento:** El MANOVA permite examinar cómo diferentes variables ambientales afectan simultáneamente múltiples aspectos del crecimiento de las plantas. Por ejemplo, se pueden analizar al mismo tiempo la altura, el diámetro del tallo, el área foliar y la biomasa en relación con factores como la temperatura, la humedad y la disponibilidad de nutrientes (Scheiner, 2001).
- **Consideración de las interrelaciones entre variables de crecimiento:** A diferencia de realizar múltiples ANOVA separados, el MANOVA tiene en cuenta las correlaciones entre las diferentes medidas de crecimiento. Esto es importante porque los diversos aspectos del crecimiento de las plantas suelen estar interrelacionados (Johnson & Wichern, 2007).

- **Control del error Tipo I:** Al analizar múltiples variables de crecimiento en una sola prueba, el MANOVA ayuda a controlar la tasa de error Tipo I que podría inflarse si se realizaran múltiples pruebas separadas para cada variable de crecimiento (Huberty & Morris, 1989).
- **Detección de efectos multivariados:** El MANOVA puede revelar efectos de las variables ambientales sobre el crecimiento de las plantas que podrían no ser evidentes cuando se examina cada variable de crecimiento por separado. Esto es particularmente útil cuando los efectos de los factores ambientales son sutiles o complejos (Scheiner & Gurevitch, 2001).
- **Análisis de interacciones complejas:** El MANOVA permite examinar cómo las interacciones entre diferentes variables ambientales afectan múltiples aspectos del crecimiento de las plantas simultáneamente. Por ejemplo, se puede analizar cómo la interacción entre temperatura y disponibilidad de agua afecta conjuntamente la altura, la biomasa y el área foliar (Tabachnick & Fidell, 2019).
- **Comparación de respuestas de crecimiento entre especies o variedades:** El MANOVA puede ser útil para comparar cómo diferentes especies o variedades de plantas responden a las mismas variables ambientales en términos de múltiples aspectos de crecimiento (Poorter et al., 2012).
- **Análisis de cambios en el crecimiento a lo largo del tiempo:** Utilizando un diseño de medidas repetidas, el MANOVA puede ayudar a analizar cómo los patrones de crecimiento de múltiples variables cambian a lo largo del tiempo en respuesta a factores ambientales (von Ende, 2001).

Tabla 2
Ejemplos de uso de MANOVA y ANOVA

Título	Resumen	Detalles
<p>Drought and Heat Stress Injury to Two Cool-Season Turfgrasses in Relation to Antioxidant Metabolism and Lipid Peroxidation</p>	<p>La sequía y las altas temperaturas son dos factores importantes que limitan el crecimiento del césped de estación fría durante el verano en muchas áreas. El objetivo del estudio fue examinar si los efectos adversos de la sequía y el calor solos o en combinación en la festuca alta (<i>Festuca arundinacea</i> L.) y el pasto azul de Kentucky (<i>Poa pratensis</i> L.) involucran estrés oxidativo. Los pastos fueron expuestos a la sequía (suspensión del riego), calor (35 °C/30 °C) y los estreses combinados durante 30 días en cámaras de crecimiento. La calidad del césped (TQ), el contenido relativo de agua de las hojas (RWC) y el contenido de clorofila (Chl) disminuyeron con la sequía prolongada, el calor y el estrés combinado para ambas especies, pero la gravedad de la disminución varió con el tipo de estrés y la duración. Aumentos transitorios en las actividades de superóxido dismutasa (SOD), ascorbato peroxidasa (AP) y glutatión reductasa (GR) ocurrieron a los 6 o 12 días de sequía y los estreses</p>	<p>Link de alojamiento: https://access.onlinelibrary.wiley.com/doi/10.2135/cropsci2001.412436x</p> <p>Autores y año: Yiwei Jiang, Bingru Huang</p> <p>Año de publicación (2001)</p>

combinados en ambas especies; Sin embargo, las actividades de las tres enzimas disminuyeron con períodos prolongados de sequía y los estreses combinados. La actividad de SOD no se vio afectada por el estrés térmico solo. Las actividades de AP y GR se redujeron después de 18 días de estrés térmico para ambas especies, pero las reducciones fueron menores que bajo los estreses combinados. La actividad de la catalasa (CAT) continuó disminuyendo por debajo del nivel de control, comenzando a los 12 días para las plantas estresadas por sequía o calor y a los 6 días para las plantas expuestas a los estreses combinados. La peroxidación lipídica ocurrió después de 18 días de estrés en ambas especies, como lo indica el aumento del contenido de malondialdehído (MDA). Los resultados sugirieron que los daños por sequía, calor o los estreses combinados tanto a la festuca alta como al pasto azul de Kentucky, como se manifiesta por disminuciones en TQ, RWC y Chl, podrían estar asociados con una disminución en las actividades de las enzimas antioxidantes y un aumento en la peroxidación lipídica de la membrana.

Nota. En esta tabla se muestran artículos en los cuales se emplearon las diferentes técnicas mencionadas para poder llegar a resultados de investigación

Técnicas para Relleno de Datos Vacíos

Las técnicas de relleno de datos vacíos, también conocidas como métodos de imputación, son estrategias fundamentales en el campo de la estadística y la ciencia de datos para abordar el problema de los valores faltantes en conjuntos de datos. Estos métodos son cruciales para mantener la integridad de los análisis estadísticos, evitar sesgos en los resultados y maximizar la utilidad de los datos disponibles (Little & Rubin, 2019).

Tipos de datos faltantes

- **MCAR (Missing Completely At Random):** Los datos faltan de manera completamente aleatoria.
- **MAR (Missing At Random):** La probabilidad de que falten datos depende de los datos observados, pero no de los datos faltantes.
- **MNAR (Missing Not At Random):** La probabilidad de que falten datos depende de los propios valores faltantes.

Donde según Enders “La elección de la técnica de imputación depende en gran porcentaje de acierto del tipo de datos faltantes a los que se enfrente” (Enders, 2010).

Para esto se define algunas de las técnicas de relleno de datos faltantes que son más relevantes como las siguientes:

- **Eliminación de Casos Completos (Listwise Deletion):** Esta técnica implica eliminar todas las observaciones que contienen uno o más valores faltantes en cualquiera de las variables de interés.

Pasos del funcionamiento de la técnica

- a. Se identifica cada fila del conjunto de datos que tiene al menos un valor faltante.
- b. Estas filas se eliminan por completo del conjunto de datos.
- c. El análisis se realiza únicamente con las filas restantes que tienen datos completos.

Para la aplicación de esta técnica de relleno de datos se comprende desde las siguientes aplicaciones

- d. Se utiliza cuando la proporción de datos faltantes es pequeña (generalmente menos del 5% de las observaciones).
 - e. Es más apropiada cuando se asume que los datos faltan completamente al azar (MCAR).
 - f. Comúnmente aplicada en análisis de regresión, ANOVA, y otros métodos estadísticos que requieren conjuntos de datos completos.
- **Imputación por Media/Mediana/Moda:** Esta técnica reemplaza los valores faltantes con una medida de tendencia central (media, mediana o moda) de la variable en cuestión, calculada a partir de los valores no faltantes.
 - **Imputación por Regresión:** Esta técnica utiliza un modelo de regresión para predecir los valores faltantes basándose en las relaciones entre la variable con datos faltantes y otras variables del conjunto de datos.

Esta técnica que aplica la regresión funciona de la siguiente forma:

- a. Se identifica la variable con datos faltantes (variable dependiente).
- b. Se seleccionan otras variables completas como predictores (variables independientes).
- c. Se construye un modelo de regresión utilizando los casos completos.
- d. Se utiliza el modelo para predecir los valores faltantes.
- e. Los valores predichos se utilizan para rellenar los datos faltantes.

Para una aplicación adecuada de esta técnica es viable aplicarla cuando existe una fuerte relación entre la variable que se evalúa con los datos faltantes y otras variables del conjunto de datos, además de que puede ser lineal o no lineal dependiendo del origen y la relación de las variables que se tenga

- **Imputación por k Vecinos Más Cercanos (KNN):** Esta técnica utiliza la similitud entre observaciones para estimar los valores faltantes, basándose en los valores de los k casos más similares o "vecinos más cercanos".

Para implementar esta técnica se realiza el siguiente proceso

- a. Se calculan medidas de distancia (como la distancia euclidiana) entre esta observación y todas las demás en el conjunto de datos, utilizando las variables completas.
- b. Se identifican los k casos más similares (vecinos más cercanos) basados en estas distancias.
- c. Se utiliza la información de estos k vecinos para imputar el valor faltante (por ejemplo, tomando la media o la mediana de los valores de los vecinos).

Esta técnica es útil cuando existen patrones en la captura de los datos, teniendo una mejor respuesta cuando las relaciones entre las variables son complejas de interpretar o son no lineales, de esta forma se puede aplicar tanto a variables que son numéricas y categóricas.

El Análisis Descriptivo en el Crecimiento de las Plantas

El análisis descriptivo es una herramienta fundamental para comprender el crecimiento de las plantas en relación con las variables ambientales. Este enfoque estadístico permite a los investigadores resumir, organizar y presentar datos de manera que faciliten la interpretación de patrones y tendencias en el desarrollo vegetal (Gómez-Hernández & Martínez-Soto, 2020).

Una de las principales formas en que el análisis descriptivo contribuye a este campo es a través de la visualización de patrones. Por ejemplo, mediante el uso de gráficos de dispersión, los científicos pueden observar la relación entre la temperatura promedio diaria y la tasa de crecimiento de las plantas, lo que ayuda a identificar rangos óptimos de temperatura para diferentes especies (Huang et al., 2021).

La comparación entre grupos es otra aplicación valiosa del análisis descriptivo. Pérez-López y Rodríguez-Gómez (2019) demostraron cómo los diagramas de caja pueden comparar eficazmente el crecimiento de plantas expuestas a diferentes intensidades de luz, revelando cómo la disponibilidad de luz afecta el desarrollo de distintas especies.

En cuanto a la precipitación, Smith y Johnson (2022) sugieren el uso de histogramas o gráficos de barras para ilustrar cómo varía el crecimiento de las plantas en función de diferentes niveles de precipitación. Esta técnica permite determinar los requerimientos

hídricos específicos de cada especie, información crucial para la gestión de cultivos y la conservación de plantas.

La relación entre los nutrientes del suelo y el crecimiento vegetal también puede ser analizada mediante técnicas descriptivas. Zhang et al. (2023) proponen el uso de gráficos de líneas para mostrar cómo cambia el crecimiento de las plantas a lo largo del tiempo en suelos con diferentes niveles de nutrientes. Este enfoque es particularmente útil para identificar deficiencias o excesos nutricionales que puedan afectar el desarrollo de las plantas.

El análisis descriptivo no solo proporciona una visión general del crecimiento de las plantas, sino que también sienta las bases para análisis más complejos. Al identificar tendencias y relaciones preliminares, los investigadores pueden formular hipótesis más precisas y diseñar experimentos más enfocados (Huang et al., 2021).

1.2. Línea de Investigación

El proyecto hace parte de la línea de investigación “Aprendizaje, conocimiento, tecnologías, comunicación y digitalización (Universidad de Cundinamarca, 2019)” de la universidad de Cundinamarca, por medio del proyecto se busca poder brindar herramientas que permitan tomar decisiones basadas en datos, basadas en estas decisiones poder generar estrategias que busquen ayudar a mejorar la producción.

1.3. Planteamiento del Problema y Pregunta de Investigación

1.3.1. Planteamiento del Problema

El cultivo de arándanos Biloxi enfrenta desafíos en la gestión del crecimiento y la sostenibilidad ambiental. Adaptar este cultivo a un ambiente controlado con recirculación

de agua es complejo debido a limitaciones de espacio, recursos y conocimientos de agricultura de precisión en la región.

Un reto adicional es la variabilidad de condiciones ambientales dentro del cultivo. Algunas áreas reciben luz solar insuficiente, mientras que, en otros obstáculos dificultan la medición precisa de variables ambientales clave. Esta heterogeneidad puede resultar en un desarrollo desigual de las plantas y decisiones de manejo subóptimas.

Para abordar estos desafíos, es crucial desarrollar enfoques de manejo más precisos y controlados. Esto implica implementar buenas prácticas agrícolas, controlando el uso de recursos y establecer un sistema de monitoreo ambiental. Una red de sensores estratégicamente ubicados permitiría comprender y manejar las variaciones micro climáticas, ajustando parámetros como riego, fertilización e iluminación de forma localizada.

Este tipo de herramientas permitirá tener una mejor perspectiva del cultivo, además de esto llevar al agricultor a tomar decisiones basadas en datos descriptivos de las plantas acercándose más a una agricultura de precisión

Por lo que la pregunta problema que se aborda es:

1.3.2. Pregunta de Investigación

¿Cómo se puede describir el crecimiento de los arándanos Biloxi en un sistema cerrado y abierto de recirculación de agua teniendo en cuenta las variables ambientales de su entorno?

1.4. *Objetivo general y objetivos específicos*

1.4.1. *Objetivo general*

Desarrollar un modelo matemático para describir el crecimiento de arándanos mediante comparación de variables ambientales en un sistema cerrado y abierto.

1.4.2. *Objetivos específicos*

1. Desarrollar un módulo de software para la visualización del modelo matemático en un servicio Web.
2. Integrar los datos obtenidos sobre el entorno y el crecimiento de la planta de arándanos Biloxi.
3. Desarrollar un modelo matemático para describir el crecimiento de la planta de arándano Biloxi usando herramientas informáticas.
4. Realizar pruebas funcionales y no funcionales del módulo de software.

1.5. *Alcance e Impacto del Proyecto*

El objetivo de este proyecto fue desarrollar un modelo matemático descriptivo que explicara el crecimiento de la planta de arándanos Biloxi por medio de una comparación en función de las variables ambientales que fueron medidas con sensores y métodos de análisis. De esta forma, este modelo matemático permitirá a los productores de arándanos evaluar y comparar el crecimiento de las plantas de arándanos en sistemas hidropónicos cerrados, lo que facilitará la toma de decisiones basadas en evidencia para su producción. Asimismo, la aplicación de este modelo podría mejorar la calidad de la producción de arándanos, ofreciendo un producto más sustentable y rentable para los consumidores. Por ende, se contribuyó a la agricultura inteligente o agricultura de precisión en la región.

Las prácticas agronómicas tradicionales como el uso de fertilizantes, pesticidas y herbicidas tienen un efecto adverso sobre el medio ambiente, a pesar de que los cultivos de arándanos colombianos son rentables en el mercado global (Losada Pardo, 2021). Muchos productores están implementando prácticas agrícolas sostenibles como el control biológico de plagas y la reducción del uso de productos químicos para disminuir los efectos adversos de los productos químicos en el medio ambiente. (Nueva resolución para el registro y uso de plaguicidas químicos y bioinsumos en los cultivos de arándanos | ICA - Instituto Colombiano Agropecuario, s. f). Otra opción prometedora es cultivar arándanos en espacios cerrados, pero esto plantea dificultades para desarrollar nuevos métodos de control y seguimiento específicos de estos cultivos. Como resultado, es esencial investigar nuevas técnicas y tecnologías que permitan a los productores de arándanos cultivar de manera sostenible en entornos regulados.

Por su énfasis en la sostenibilidad agrícola y la innovación tecnológica, esta propuesta está estrechamente relacionada con los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas. Este proyecto contribuye específicamente al ODS 9, que pretende fomentar la innovación y el uso de tecnologías de vanguardia para elevar la sostenibilidad de la industria. Se prevé que los productores locales de arándanos reciban asistencia en la producción como resultado del desarrollo de este modelo matemático, lo que les ayudará a ser más competitivos en el mercado.

El proyecto también se relaciona con el ODS número 12, que tiene como objetivo promover la producción y el consumo sostenibles a través de prácticas agrícolas y de gestión de recursos más responsables. Se espera que los productores locales de arándanos podrán reducir los recursos naturales, como agua y energía, y producir menos residuos y

productos químicos gracias al uso de la tecnología. Esto ayudará a que la producción de arándanos de la región deje una huella ecológica menor.

Un modelo matemático orientado a describir el comportamiento de la planta de arándanos puede ofrecer beneficios significativos para una producción sostenible, con impactos a corto, mediano y largo plazo. A corto plazo, mejoraría el cultivo mediante variables medibles, permitiendo tomar medidas correctivas para gestionar condiciones no ideales. A mediano plazo, la solución matemática impulsaría una agricultura de precisión más sustentable. A largo plazo, proporcionaría una comprensión más clara de un ambiente para el correcto desarrollo de la planta, contribuyendo a sistemas alimentarios sostenibles. Una vez adaptado, este modelo permitiría describir el comportamiento del cultivo bajo diversas condiciones ambientales y prácticas agrícolas, facilitando un manejo más informado y una toma de decisiones más precisa. Este enfoque científico para mejorar cultivos intensivos es crucial en la transición hacia la sostenibilidad, alineándose con los Objetivos de Desarrollo Sostenible de la ONU y sentando las bases para una producción y consumo más sostenibles.

1.6. Metodología

Este proyecto adopto un enfoque comparativo y descriptivo, centrado en el análisis estadístico de datos sobre el crecimiento y comportamiento de las plantas de arándanos. Se recopilaron datos sobre variables clave como altura de la planta, pH, temperatura, etc. Estos datos fueron analizados utilizando técnicas estadísticas como correlación, regresión y análisis de varianza, representaciones graficas que permita ver la comparación y entre otros.

El objetivo se centró en la idea de desarrollar un modelo matemático que comparara la variable independiente (tiempo) y la variable dependiente del crecimiento y

comportamiento de la planta, además de que pudiera brindar una descripción por este medio. Se exploraron por medio del estado del arte modelos ya implementados en otros estudios y sus diferentes enfoques para abordar el tema de manera más asertiva lo que permitió brindar una descripción del trabajo que se esté realizando.

Para la programación y desarrollo del modelo matemático y el módulo de software, se implementó la metodología ágil SCRUM. Se organizaron sprints en ciclos de trabajo durante los cuales el equipo completo tareas específicas.

La validación del modelo se realizó comparando el análisis con datos reales de crecimientos recolectados en campo. Se busco minimizar el error cuadrático medio entre los valores medidos y observados. El modelo fue refinado de manera iterativa, incorporando nuevos datos y técnicas estadísticas más adecuadas.

Una vez fue obtenido un modelo matemático adaptado, este se utilizó para realizar descripciones. Por ejemplo, describir el comportamiento bajo diferentes condiciones ambientales o prácticas de cultivo. Esto permitió tener una idea más clara sobre el manejo del cultivo y la toma de decisiones. La metodología cuantitativa propuesta, enfocada en modelado matemático, análisis estadístico y SCRUM, permitió una mejor comprensión de la biología y fisiología del crecimiento de las plantas de arándanos.

1.7. Marcos de Referencia

1.7.1. Marco teórico

Para el desarrollo del proyecto es fundamental tener presentes los conceptos como la agricultura de precisión, modelos que hay de agricultura y el fuerte que son los modelos matemáticos, en el desarrollo del marco teórico se presentaran algunos modelos que se han

destacado en el estado del arte para poder conocer un poco de cada uno de estos, también se complementaran con técnicas de análisis de datos como análisis de varianza, análisis descriptivos, etc.

Agricultura de precisión

La agricultura de precisión es un enfoque de gestión agrícola que utiliza tecnologías de información y comunicación para optimizar la producción de cultivos y el manejo de recursos. Según Pierce y Nowak (1999), la agricultura de precisión se define como "la aplicación de tecnologías y principios para manejar la variabilidad espacial y temporal asociada con todos los aspectos de la producción agrícola con el propósito de mejorar el rendimiento del cultivo y la calidad ambiental" (p. 1).

Figura 1
Agricultura de Precisión



Nota. Imagen tomada de <https://www.solinftec.com/es-la/blog/agricultura-de-precision/>.

Para poder tener una implementación de una agricultura de precisión se deben tener varios componentes que son claves para el correcto funcionamiento de esta técnica de agricultura.

- **Sistemas de Posicionamiento Global (GPS):** Permiten la localización precisa de máquinas agrícolas y áreas específicas del campo (Stafford, 2000).
- **Sistemas de Información Geográfica (SIG):** Ayudan a almacenar, analizar y visualizar datos espaciales del campo (Zhang et al., 2002).
- **Sensores Remotos:** Incluyen imágenes satelitales y drones para monitorear el estado de los cultivos y las condiciones del suelo (Mulla, 2013).
- **Sensores in situ:** Miden variables como humedad del suelo, temperatura y niveles de nutrientes en tiempo real (Gebbers & Adamchuk, 2010).
- **Maquinaria de Tasa Variable:** Permite la aplicación precisa de insumos como fertilizantes y pesticidas basada en las necesidades específicas de cada área del campo (Robert, 2002).

Estos son algunos componentes claves para poder implementar una agricultura de precisión, además de todo lo anteriormente descrito se quiere también resaltar de como esta técnica contribuye al control de los recursos naturales, donde se encontró lo siguiente en donde se presentan algunos factores en donde la agricultura de precisión ayuda a tener un mejor control de los recursos naturales.

- **Uso Eficiente del Agua:** Mediante sistemas de riego de precisión que aplican agua solo donde y cuando es necesario, reduciendo el desperdicio (Sadler et al., 2005).

- **Optimización de Fertilizantes:** La aplicación de tasas variables de fertilizantes basada en mapas de nutrientes del suelo reduce la sobreaplicación y la lixiviación (Bongiovanni & Lowenberg-DeBoer, 2004).
- **Reducción de Pesticidas:** El uso de mapas de infestación de plagas permite una aplicación más precisa y reducida de pesticidas (Gerhards & Oebel, 2006).
- **Conservación del Suelo:** Las prácticas de labranza de precisión ayudan a reducir la erosión y mejorar la salud del suelo (Lal, 2020).
- **Mitigación del Cambio Climático:** La agricultura de precisión puede reducir las emisiones de gases de efecto invernadero al optimizar el uso de insumos y mejorar la eficiencia energética (Balafoutis et al., 2017).

Para finalizar la agricultura de precisión representa un cambio paradigmático en la gestión agrícola, permitiendo un uso más eficiente y sostenible de los recursos naturales. Como señalan Gebbers y Adamchuk (2010), "la agricultura de precisión promete aumentar la productividad mientras reduce los costos y minimiza los impactos ambientales" (p. 828).

Como se puede evidenciar en este apartado del marco teórico, la agricultura de precisión juega un papel fundamental para el desarrollo de soluciones tecnológicas que permiten tomar decisiones basadas en datos concretos. La implementación de sistemas automatizados de monitoreo y control representa una evolución significativa en la forma en que entendemos y manejamos los cultivos, ya que permite obtener información precisa y en tiempo real sobre las condiciones que se presentan el desarrollo de las plantas y de cómo estas se comportan en diferentes ambientes.

Esta nueva forma de agricultura tecnificada no solo mejora nuestra comprensión del crecimiento y comportamiento de las plantas en diferentes ambientes, sino que también

mejora el uso de los recursos mediante el uso estratégico de sensores y actuadores. Los sensores proporcionan datos vitales sobre variables ambientales, mientras que los actuadores permiten realizar ajustes precisos para mantener las condiciones adecuadas para cada cultivo. Este enfoque integral de la agricultura moderna demuestra cómo la tecnología se ha convertido en una herramienta necesaria para mejorar la producción agrícola y poder tener una mejor sostenibilidad de los cultivos en diversos entornos controlados.

Modelos de agricultura

La agricultura ha evolucionado significativamente a lo largo de la historia, adaptándose a las necesidades cambiantes de la sociedad y a los avances tecnológicos. Este marco teórico explora los principales modelos de agricultura que se han desarrollado y se practican en la actualidad.

- **Agricultura Tradicional:** La agricultura tradicional se caracteriza por métodos de cultivo heredados a través de generaciones, con un uso mínimo de tecnología moderna (Altieri, 2004). Este modelo se basa en el conocimiento local y las prácticas culturales específicas de cada región.

Figura 2
Agricultura tradicional



Nota. Tema cultivo de tierras sin tecnificación y artesanal. Imagen tomada de <https://encolombia.com/economia/agroindustria/agricultura-tradicional/>

Características principales

- Uso de variedades locales de cultivos
- Policultivos y rotación de cultivos
- Dependencia de recursos locales y reciclaje de nutrientes
- Manejo integrado de plagas basado en métodos naturales

Gliessman (2015) argumenta que la agricultura tradicional a menudo es más sostenible y resiliente que los sistemas agrícolas industriales modernos, debido a su adaptación a las condiciones locales y su menor dependencia de insumos externos.

- **Agricultura Convencional o Industrial:** La agricultura convencional, también conocida como industrial, se desarrolló durante la Revolución Verde y se caracteriza por un alto uso de insumos externos y tecnología para maximizar la producción (Pingali, 2012).

Figura 3
Agricultura industrial.



Nota. Tema Industria Agrícola.

Características principales

- Monocultivos a gran escala
- Alto uso de fertilizantes químicos y pesticidas
- Mecanización intensiva
- Uso de variedades de cultivos de alto rendimiento

Aunque este modelo ha aumentado significativamente la producción global de alimentos, Foley et al. (2011) señalan que también ha llevado a problemas ambientales como la degradación del suelo, la contaminación del agua y la pérdida de biodiversidad.

- **Agricultura Orgánica:** La agricultura orgánica es un sistema de producción que mantiene y mejora la salud de los suelos, los ecosistemas y las personas, evitando el uso de fertilizantes y pesticidas sintéticos (Reganold & Wachter, 2016).

Características principales

- Uso de compost y abonos orgánicos

- Control biológico de plagas
- Rotación de cultivos y diversidad de especies
- Prohibición de OGMs

Seufert et al. (2012) encontraron que, aunque los rendimientos de la agricultura orgánica son generalmente más bajos que los de la agricultura convencional, los beneficios ambientales y de salud pueden compensar esta diferencia.

- **Agricultura de Conservación:** La agricultura de conservación se enfoca en mejorar y sostener la productividad agrícola, aumentar los beneficios y la seguridad alimentaria mientras se preservan y mejoran los recursos y el medio ambiente (FAO, 2021).

Características principales

- Mínima perturbación del suelo (labranza cero o reducida)
- Cobertura permanente del suelo
- Diversificación de especies de cultivos

Hobbs et al. (2008) destacan que este modelo puede mejorar la calidad del suelo, reducir la erosión y aumentar la eficiencia en el uso del agua.

- **Agricultura vertical:** La agricultura vertical es un método de producción de alimentos en capas apiladas verticalmente, a menudo en ambientes controlados como edificios o contenedores (Despommier, 2010).

Figura 4
Agricultura Vertical



Nota. Tema: “El futuro de la alimentación se llama agricultura vertical”. Imagen tomada de <https://hazrevista.org/rsc/2018/08/el-futuro-de-la-alimentacion-se-llama-agricultura-vertical/>

Características principales

- Cultivo en ambientes cerrados y controlados
- Uso de iluminación artificial y sistemas hidropónicos
- Alta eficiencia en el uso del espacio y el agua
- Producción durante todo el año independiente del clima exterior

Al-Chalabi (2015) sugiere que la agricultura vertical puede ser una solución para la producción de alimentos en áreas urbanas con espacio limitado, aunque los costos energéticos pueden ser un desafío.

De este apartado se concluye que en la actualidad coexisten diversos modelos de agricultura, cada uno con características distintivas que los hacen únicos en su aplicación y desarrollo. Entre estos modelos destaca especialmente la agricultura vertical, considerada como el futuro de los cultivos por su capacidad de adaptarse a entornos urbanos y su

estrecha relación con la agricultura de precisión. Esta innovadora aproximación representa una evolución significativa respecto a los métodos tradicionales e industriales, demostrando cómo la integración de tecnología puede transformar radicalmente la manera en que producimos nuestros alimentos.

La implementación de herramientas avanzadas, como un modelo matemático, permite tener una mejor comprensión del comportamiento de las plantas en diferentes condiciones de cultivos, especialmente en sistemas de recirculación de agua. Este modelo, junto con el control preciso de variables específicas en ambientes controlados, no solo permite tener un control sobre el desarrollo de las plantas, sino que también contribuyen a una producción más eficiente. Como resultado, se logra una reducción significativa en la pérdida de recursos valiosos como fertilizantes y se maximiza la calidad del producto final, representando así un avance sustancial hacia una agricultura más sostenible y productiva.

Modelos matemáticos

Modelos estadísticos: Estos tipos de modelos generalmente son conocidos como: estadísticos (Kisiel et al, 2023), descriptivos (Ruiz et al. 2011) o empíricos (Domijan et al. 2006; Hernandez et al, 2009), pero para normalizar una referencia a esta metodología de modelado se van a referir a ellos a partir de ahora cómo modelos estadísticos.

Los modelos estadísticos incluyen diferentes técnicas que usan grandes bases de datos que, por lo general describen datos reales dentro de un sistema de forma empírica, siendo estos datos variables que interactúan y evolucionan entre sí en el tiempo (Bravo et al. 2019). Algunas formas de tratar con este tipo de variables son basadas en la regresión, cómo: regresión lineal simple, regresión lineal Estos tipos de modelos generalmente son conocidos como: estadísticos (Kisiel et al, 2023), descriptivos (Ruiz et al. 2011) o

empíricos (Domijan et al. 2006; Hernandez et al, 2009), pero para normalizar una referencia a esta metodología de modelado se van a referir a ellos a partir de ahora cómo modelos estadísticos.

Modelos comparativos y descriptivos: Los modelos comparativos y descriptivos son herramientas estadísticas utilizadas en diversas áreas del conocimiento para analizar y entender las características y relaciones entre variables. Estos modelos permiten realizar comparaciones sistemáticas y describir patrones y tendencias en los datos.

Los modelos comparativos se enfocan en evaluar si existen diferencias significativas entre grupos o tratamientos respecto a una o más variables de interés. Como señalan Manzano et al. (2019), "los modelos comparativos buscan determinar si las diferencias observadas entre grupos son estadísticamente significativas o si, por el contrario, podrían deberse al azar" (p. 49). Algunos ejemplos de modelos comparativos incluyen el análisis de varianza (ANOVA), las pruebas de comparaciones múltiples y el análisis discriminante.

Por otro lado, los modelos descriptivos tienen como objetivo principal resumir y caracterizar la distribución de una o más variables en una población o muestra. Según Leandro et al. (2019), "los modelos descriptivos permiten explorar y comprender la estructura subyacente de los datos, identificar patrones y relaciones entre variables, y proporcionar una representación visual efectiva de la información" (p. 23). Algunas técnicas descriptivas comúnmente utilizadas son las medidas de tendencia central (media, mediana, moda), las medidas de dispersión (varianza, desviación estándar), los gráficos (histogramas, diagramas de caja, gráficos de dispersión) y el análisis de correlación.

Tanto los modelos comparativos como los descriptivos son esenciales en el proceso de investigación y análisis de datos. Como afirman Martínez-Herrera et al. (2019), "la combinación de modelos comparativos y descriptivos proporciona una comprensión más profunda de los fenómenos estudiados, permitiendo no solo identificar diferencias significativas, sino también explorar las características y relaciones subyacentes de las variables" (p. 78).

Modelos aditivos generalizados: Los Modelos Aditivos Generalizados (GAM, por sus siglas en inglés) son una extensión flexible de los modelos lineales generalizados (GLM) introducidos por Hastie y Tibshirani (1990). Los GAM permiten modelar relaciones no lineales entre las variables predictoras y la variable respuesta, manteniendo la interpretabilidad de los modelos lineales.

$$g(E[Y]) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_k(x_k)$$

Donde:

- **Y** Es la variable respuesta
- **g()** es la función enlace
- **E[Y]** es el valor esperado de **Y**
- **β_0** es el intercepto
- **$f_1(x_1)$** son funciones suaves (no paramétricas) de las variables predictoras X_i

La implementación conjunta de modelos matemáticos estadísticos, comparativos y descriptivos, junto con el modelo GAM (Modelo Aditivo Generalizado), representa una buena herramienta para la comprensión profunda del crecimiento vegetal en ambientes controlados. La combinación entre estos tres enfoques, respaldada por técnicas analíticas

como MANOVA, ANOVA y análisis discriminante, permite obtener una visión bastante fuerte de cómo las diferentes variables ambientales influyen en el desarrollo de las plantas. Estas metodologías estadísticas no solo facilitan la identificación de patrones de crecimiento específicos, sino que también posibilitan la detección de relaciones complejas entre múltiples variables que podrían pasar desapercibidas.

La aplicación de estas herramientas estadísticas avanzadas pretende ayudar a la toma de decisiones en la agricultura controlada, permitiendo ajustes más precisos y fundamentados en los parámetros de cultivo. El análisis multivariado a través de MANOVA y el análisis discriminante, combinado con la flexibilidad del modelo GAM, proporciona a los agricultores y científicos la capacidad de describir de forma detallada el crecimiento de las plantas en diversos ambientes artificiales. Esta aproximación basada en datos no solo mejora la producción, sino que también permite desarrollar protocolos de cultivo más adecuados y adaptables a diferentes condiciones ambientales, marcando así un avance significativo en la agricultura de precisión y el cultivo en ambientes controlados.

1.7.2. Marco legal

En este marco legal se presentan normativas que se deben tener presentes en Colombia para la agricultura y la agricultura de precisión que son las que se presentan a continuación

Normas para la agricultura en Colombia

- **Resolución 1056 de 2022 del Instituto Colombiano Agropecuario (ICA)**

Esta resolución, emitida por el ICA, regula el uso de productos fitosanitarios en Colombia. Los productos fitosanitarios incluyen pesticidas, herbicidas, fungicidas y otros

químicos utilizados para proteger los cultivos de plagas y enfermedades. El objetivo principal de esta norma es asegurar que los productos que se utilizan en la agricultura cumplan con altos estándares de seguridad para la salud humana y el medio ambiente. Para ello, la resolución establece los requisitos que deben cumplir los fabricantes e importadores de estos productos en cuanto a su registro, comercialización y manejo.

Además, la norma obliga a los agricultores a aplicar estos productos de acuerdo con las dosis y condiciones recomendadas por el fabricante, evitando su uso excesivo o inadecuado que podría generar problemas de contaminación en los suelos, aguas o alimentos. El ICA también establece controles sobre la capacitación y certificación de los aplicadores de fitosanitarios, promoviendo el uso de equipos de protección personal y prácticas que minimicen el impacto ambiental.

Esta resolución es esencial para garantizar que la agricultura en Colombia sea sostenible y que no se comprometa la salud pública ni el equilibrio de los ecosistemas.

- **Ley 1876 de 2017 - Sistema Nacional de Innovación Agropecuaria**

La Ley 1876 de 2017 crea el Sistema Nacional de Innovación Agropecuaria (SNIA), cuyo propósito es fortalecer el sector agrícola mediante la investigación, el desarrollo y la adopción de nuevas tecnologías. Este sistema busca integrar a los actores clave del sector agropecuario, como los agricultores, investigadores, entidades del gobierno y el sector privado, para mejorar la productividad y la sostenibilidad de la agricultura en Colombia.

La ley establece que el SNIA debe enfocarse en cuatro áreas principales:

- **Investigación y Desarrollo (I+D):** Promover la investigación científica y tecnológica para el desarrollo de nuevas técnicas agrícolas que optimicen la producción de alimentos y reduzcan el impacto ambiental.
- **Transferencia de Tecnología:** Facilitar la adopción de tecnologías emergentes por parte de los agricultores, a través de programas de capacitación y asistencia técnica.
- **Innovación en Servicios:** Fomentar la creación de nuevos productos y servicios en el sector agrícola que aumenten el valor añadido de los productos agropecuarios.
- **Articulación del Conocimiento:** Generar sinergias entre los actores del sector para el intercambio de conocimientos y experiencias.

Esta ley es clave para modernizar la agricultura colombiana y asegurar que los avances tecnológicos beneficien a todo el sector agropecuario, especialmente a los pequeños y medianos agricultores, quienes a menudo carecen de acceso a tecnologías de vanguardia.

- **Resolución 180594 de 2019 - Reglamento Técnico de Fertilizantes**

La Resolución 180594 de 2019, expedida por el Ministerio de Agricultura y Desarrollo Rural, regula la fabricación, comercialización y uso de fertilizantes en Colombia. El principal objetivo de esta resolución es asegurar que los fertilizantes utilizados en la agricultura sean de alta calidad y seguros tanto para los cultivos como para el medio ambiente y la salud de los consumidores.

Este reglamento técnico establece los requisitos que deben cumplir los productores e importadores de fertilizantes en cuanto a su composición química, etiquetado, almacenamiento y distribución. También define los procedimientos de inspección y control que deben seguir las autoridades para verificar el cumplimiento de estos requisitos.

Una parte fundamental de la resolución es que promueve el uso de fertilizantes orgánicos y biológicos, como una alternativa a los fertilizantes químicos convencionales, que pueden tener efectos negativos a largo plazo sobre los suelos y los ecosistemas acuáticos. Además, se fomenta la capacitación de los agricultores en el uso responsable de fertilizantes, para evitar problemas de sobreutilización que podrían llevar a la contaminación del agua subterránea y a la degradación de los suelos.

- **Ley 99 de 1993 - Creación del Ministerio del Medio Ambiente y Regulación de la Política Ambiental en Colombia**

La Ley 99 de 1993 es una de las normas más importantes en Colombia, ya que establece el marco regulatorio para la protección ambiental en el país. Esta ley creó el Ministerio del Medio Ambiente (hoy Ministerio de Ambiente y Desarrollo Sostenible) y regula el uso sostenible de los recursos naturales, incluidos los relacionados con la agricultura.

Esta ley impone obligaciones a los agricultores para que sus actividades no causen daño al medio ambiente y para que se implementen buenas prácticas agrícolas, como el manejo adecuado del agua, la conservación de suelos y la protección de la biodiversidad. También establece la necesidad de realizar estudios de impacto ambiental para proyectos agrícolas de gran escala.

Normas para la agricultura de precisión en Colombia

- **Ley 2046 de 2020 - Promoción de la Innovación Agropecuaria**

La Ley 2046 de 2020 tiene como objetivo promover la adopción de innovaciones tecnológicas en el sector agropecuario colombiano. Esta ley reconoce que el uso de

tecnologías avanzadas, como las que forman parte de la agricultura de precisión, es clave para mejorar la competitividad del sector y enfrentar desafíos como el cambio climático, la escasez de recursos y la necesidad de producir más alimentos para una población en crecimiento.

Entre las innovaciones que esta ley fomenta se incluyen:

- **Drones y Sensores:** Los drones permiten monitorear grandes extensiones de terreno de manera eficiente, recopilando datos en tiempo real sobre el estado de los cultivos. Los sensores, por su parte, ayudan a medir variables como la humedad del suelo, la temperatura y la calidad del aire.
- **Sistemas de Información Geográfica (SIG):** El uso de SIG en la agricultura permite hacer un seguimiento detallado de la ubicación de los cultivos y su rendimiento a lo largo del tiempo, optimizando así la gestión de los recursos.
- **Big Data y Análisis de Datos:** La recolección masiva de datos en tiempo real facilita la toma de decisiones basadas en datos, lo que permite a los agricultores ajustar sus prácticas para maximizar los rendimientos y minimizar los costos.

Esta ley es crucial para impulsar la modernización de la agricultura colombiana y para garantizar que los agricultores puedan acceder a las tecnologías más avanzadas para mejorar sus prácticas productivas.

- **Política Nacional de Agricultura de Precisión (PNAP)**

La Política Nacional de Agricultura de Precisión (PNAP) es una iniciativa del Ministerio de Agricultura y Desarrollo Rural que busca establecer un marco de referencia para la adopción de la agricultura de precisión en Colombia. Esta política no es una norma

jurídica vinculante, pero sirve como guía para la promoción de tecnologías innovadoras en el sector agrícola.

La PNAP promueve el uso de herramientas digitales, como drones, sensores y software avanzado, para mejorar la toma de decisiones en la gestión de cultivos. También busca fomentar la capacitación de los agricultores en el uso de estas tecnologías y garantizar que los pequeños productores tengan acceso a ellas, evitando la exclusión digital.

Esta política destaca la importancia de la sostenibilidad, proponiendo prácticas agrícolas que permitan una producción más eficiente en el uso de agua, energía y otros recursos, minimizando el impacto ambiental y mejorando la rentabilidad.

- **Normas ISO 22000 e ISO 14001**

Las normas internacionales ISO 22000 (Gestión de la Seguridad Alimentaria) e ISO 14001 (Gestión Ambiental) son relevantes para la agricultura de precisión en Colombia. La ISO 22000 establece los requisitos para la implementación de un sistema de gestión de la seguridad alimentaria en las empresas agrícolas, asegurando que los alimentos producidos sean seguros para el consumo. La ISO 14001, por su parte, establece un marco para la gestión responsable de los recursos naturales y la reducción del impacto ambiental.

Ambas normas son fundamentales en la agricultura de precisión, ya que el uso de tecnología avanzada permite a los agricultores cumplir con los requisitos de estas normas de manera más eficiente, mejorando la trazabilidad de los alimentos y reduciendo el uso de insumos químicos.

- **Ley 740 de 2002 - Acceso a Recursos Genéticos y Convenio sobre Diversidad Biológica**

Esta ley establece los mecanismos para regular el acceso a los recursos genéticos en Colombia, en consonancia con los compromisos del país en el Convenio sobre la Diversidad Biológica (CDB). Dado que Colombia es uno de los países más biodiversos del mundo, esta ley es crucial para proteger los recursos genéticos nativos, incluidos aquellos relacionados con la agricultura.

La ley también regula el acceso y la explotación de variedades vegetales que son importantes para la seguridad alimentaria y para el desarrollo de nuevas tecnologías agrícolas, como la mejora genética y la producción de cultivos más resilientes.

- **Resolución 970 de 2010 - Regulación sobre el Uso de Semillas Certificadas**

Esta resolución regula el uso y comercialización de semillas certificadas en Colombia, estableciendo que solo se pueden utilizar semillas registradas y aprobadas por las autoridades competentes, como el ICA. Esta norma ha sido controvertida, especialmente entre pequeños agricultores, ya que se percibe como una restricción al uso de semillas criollas o tradicionales.

- **Ley 165 de 1994 - Aprobación del Protocolo de Cartagena sobre Seguridad de la Biotecnología**

La Ley 165 de 1994 ratifica el Protocolo de Cartagena, que establece normas de bioseguridad para la manipulación, el uso y el transporte de organismos genéticamente modificados (OGM). Esta ley es esencial para garantizar que el uso de OGM en la agricultura sea seguro para el medio ambiente y la salud humana.

- **Ley 1530 de 2012 - Régimen de Regalías para la Ciencia, Tecnología e Innovación (CTI)**

Esta ley establece el Sistema General de Regalías, que incluye un fondo destinado a financiar proyectos de ciencia, tecnología e innovación en el país. Parte de estos fondos se puede destinar a proyectos agrícolas, incluidos aquellos que promueven la adopción de tecnologías de agricultura de precisión.

- **Decreto 1071 de 2015 - Decreto Único Reglamentario del Sector Agropecuario**

Este decreto consolida todas las normas del sector agropecuario en Colombia. Regula diversas actividades agrícolas, incluyendo la sanidad animal, la seguridad alimentaria, el uso de recursos naturales y el desarrollo rural. Entre otros aspectos, el decreto fomenta la adopción de prácticas agrícolas sostenibles y tecnológicamente avanzadas, que es uno de los pilares de la agricultura de precisión.

El conjunto de normas y regulaciones existentes proporciona un marco legal sólido que respalda el desarrollo de proyectos innovadores enfocados en la creación e implementación de espacios de agricultura de precisión. Estas normativas no solo promueven la implementación de tecnologías avanzadas como el uso de drones, sistemas tecnificados por medio de sensores y actuadores en ambientes de cultivos o cultivos artificiales, sino que también obligan a que estos sistemas cumplan con los estándares de calidad y seguridad necesarios. La regulación sobre el uso de productos fitosanitarios y fertilizantes asegura que los nutrientes y tratamientos aplicados en estos sistemas sean seguros y efectivos, mientras que las normas de innovación agropecuaria respaldan la investigación y desarrollo de nuevos métodos de cultivo controlado.

Por otra parte, este marco normativo impulsa la adopción de prácticas agrícolas tecnificadas que permiten un mejor control y seguimiento del desarrollo de las plantas,

buscando maximizar la producción y tener la menor pérdida del producto final. Las regulaciones sobre el manejo de recursos hídricos y la implementación de sistemas de recirculación de agua promueven la sostenibilidad y eficiencia en el uso de recursos, mientras que las normas de seguridad alimentaria garantizan que los productos cultivados en estos ambientes controlados sean aptos para el consumo humano. Este respaldo legal es fundamental para el desarrollo de proyectos que buscan mejorar la producción agrícola mediante la integración de tecnología avanzada apoyada en herramientas que permitan tomar decisiones basadas en datos como lo es el desarrollo de un modelo matemático que permita describir el crecimiento de planas de arándanos en su variedad biloxy mediante la comparación de variables ambientales en sistemas de recirculación de agua, asegurando no solo la innovación en los métodos de cultivo, sino también la calidad y seguridad de los productos obtenidos.

2. Documentación del Software

2.1. Plan de Proyecto

Figura 5

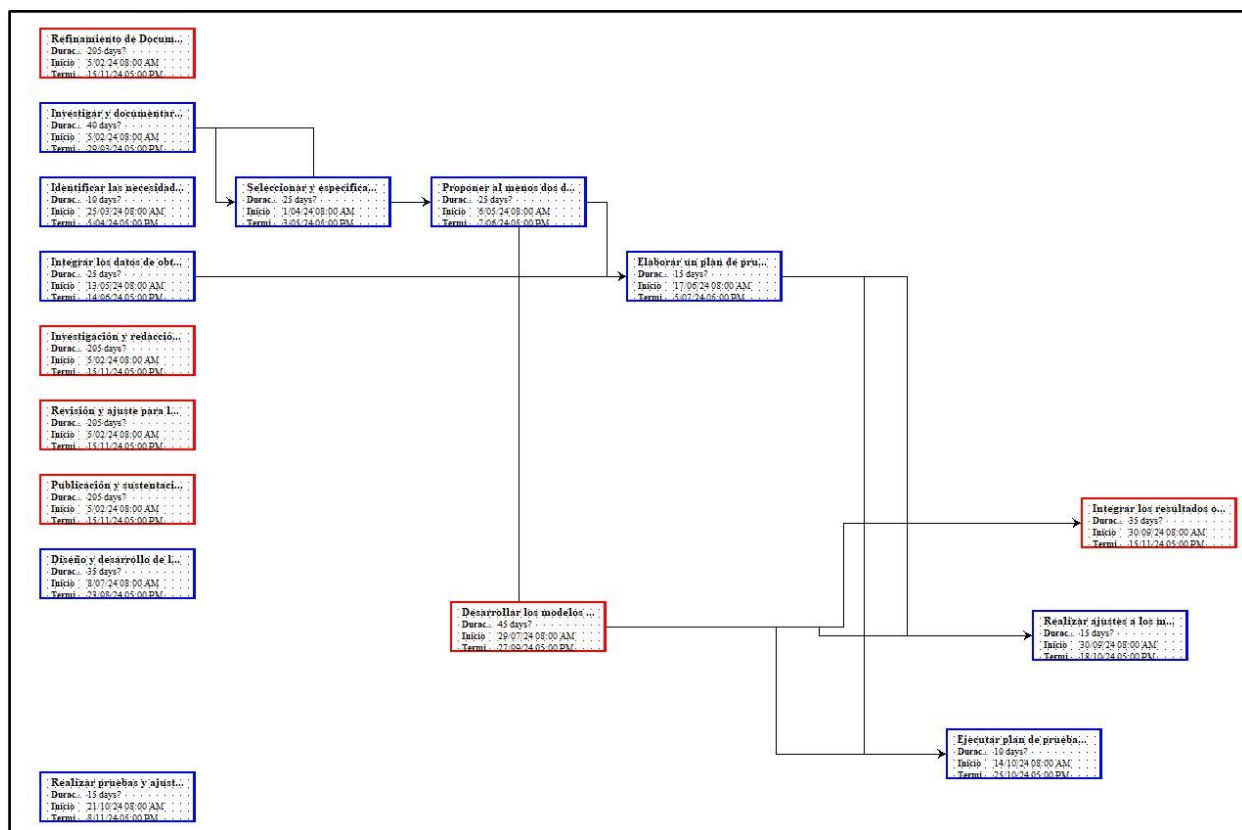
Imagen de tareas con sus respectivas sucesoras en project libre.

🕒	Nombre	Duración	Inicio	Terminado	Pred...	Nombres del Recurso
	Refinamiento de Documentación (Requerimientos, Especificaciones, etc.).	205 days?	5/02/24 08:00 AM	15/11/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
	Investigar y documentar las tecnologías y metodologías existentes en el desarrollo de modelos matemáticos y sus variantes en crecimiento de plantas.	40 days?	5/02/24 08:00 AM	29/03/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Identificar las necesidades y requerimientos del sistema y los usuarios.	10 days?	25/03/24 08:00 AM	5/04/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Seleccionar y especificar las variables descriptivas y objetivo necesarias para la consideración de un tipo de modelo matemático.	25 days?	1/04/24 08:00 AM	3/05/24 05:00 PM	2	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Proponer al menos dos diseños de modelado matemático para describir el crecimiento de arándanos en espacios reducidos, considerando variables ambientales.	25 days?	6/05/24 08:00 AM	7/06/24 05:00 PM	2;4	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Integrar los datos de obtenidos en un espacio de trabajo para generar los datasets disponibles para el desarrollo del modelo.	25 days?	13/05/24 08:00 AM	14/06/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Elaborar un plan de pruebas y comparación de los modelos matemáticos y sus diversas características	15 days?	17/06/24 08:00 AM	5/07/24 05:00 PM	5;6	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
	Investigación y redacción para la elaboración de los diferentes entregables (Artículos, ponencia y libro de grado)	205 days?	5/02/24 08:00 AM	15/11/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
	Revisión y ajuste para los diferentes entregables	205 days?	5/02/24 08:00 AM	15/11/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
	Publicación y sustentación de los artículos y ponencias	205 days?	5/02/24 08:00 AM	15/11/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Diseño y desarrollo de la interfaz gráfica para visualizar los datos y resultados.	35 days?	8/07/24 08:00 AM	23/08/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Desarrollar los modelos matemáticos propuestos usando diferentes técnicas, variables predictoras y variables objetivo.	45 days?	29/07/24 08:00 AM	27/09/24 05:00 PM	5	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Integrar los resultados obtenidos mediante el estudio realizado	35 days?	30/09/24 08:00 AM	15/11/24 05:00 PM	12	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Realizar ajustes a los modelos propuestos, identificando las falencias que presentan	15 days?	30/09/24 08:00 AM	18/10/24 05:00 PM	7;12	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Ejecutar plan de pruebas para valorar los modelos propuestos y seleccionar uno	10 days?	14/10/24 08:00 AM	25/10/24 05:00 PM	7;12	Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso
🕒	Realizar pruebas y ajustes necesarios para la integración al módulo de software	15 days?	21/10/24 08:00 AM	8/11/24 05:00 PM		Ariel Camilo Sanchez Lopez; Jhonatan Murcia Posso

Nota. Autoría propia.

En la figura 5 se presentan las tareas del proyecto organizadas en orden, junto con sus sucesoras y la asignación de tiempo en días según el cronograma establecido en el documento de investigación. Además, se destacan las tareas dependientes, es decir, aquellas que inician solo cuando se han completado sus predecesoras. Por último, se muestra la asignación de recursos encargados de ejecutar cada tarea.

Figura 6
Diagrama Gantt en formato de red.



Nota. Autoría propia.

En la figura 6 se muestran las tareas del proyecto junto con sus dependencias. Las cajas representan cada tarea, mientras que las flechas y líneas indican la relación entre ellas y el orden en que deben completarse. Además, se utiliza un código de colores para diferenciar su importancia: las tareas críticas, marcadas en rojo, son aquellas cuyo retraso afecta el cumplimiento del proyecto en los tiempos estipulados; en cambio, las tareas no críticas, representadas en azul, no impactan directamente en el cronograma general si se demoran.

Para lograr una mejor comprensión del funcionamiento del proyecto, este se ha desglosado en diferentes etapas que detallan su correcto desarrollo. Este enfoque no solo facilita

el entendimiento de sus objetivos, sino que también permite visualizar las acciones necesarias para su cumplimiento.

Tabla 3

Descripción de las tareas del proyecto

Nombre de la tarea	Tipo de tarea
Refinamiento de Documentación (Requerimientos, Especificaciones, etc.).	Documental
Investigar y documentar las tecnologías y metodologías existentes en el desarrollo de modelos matemáticos y sus variantes en crecimiento de plantas.	Documental
Identificar las necesidades y requerimientos del sistema y los usuarios.	Documental
Seleccionar y especificar las variables descriptivas y objetivo necesarias para la consideración de un tipo de modelo matemático.	Documental
Proponer al menos dos diseños de modelado matemático para describir el crecimiento de arándanos en espacios	Documental

reducidos, considerando variables ambientales.	
---	--

Integrar los datos de obtenidos en un espacio de trabajo para generar los datasets disponibles para el desarrollo del modelo.	Documental
--	------------

Elaborar un plan de pruebas y comparación de los modelos matemáticos y sus diversas características.	Documental
---	------------

Investigación y redacción para la elaboración de los diferentes entregables (Artículos, ponencia y libro de grado)	Documental
---	------------

Revisión y ajuste para los diferentes entregables	Documental y Programación
--	---------------------------

Publicación y sustentación de los artículos y ponencias	Documental
--	------------

Diseño y desarrollo de la interfaz gráfica para visualizar los datos y resultados	Programación
--	--------------

Desarrollar los modelos matemáticos propuestos usando diferentes, técnicas, variables predictoras y variables objetivo.	Documental y Programación
Integrar los resultados obtenidos mediante el estudio realizado	Documental
Realizar ajustes a los modelos propuestos, identificando las falencias que presentas	Documental y Programación
Ejecutar plan de pruebas para valorar los modelos propuestos y seleccionar uno	Documental
Realizar pruebas y ajustes necesarios para la integración al módulo de software	Programación

Nota. Autoría propia.

En la tabla se muestra las tareas del proyecto y el tipo de tarea que es cada una de estas, con el fin de poder entender de una mejor forma la estructura, desarrollo y documentación del proyecto.

2.2. Determinación de Requerimientos

Introducción: El presente documento es una especificación de requerimientos de software para el proyecto de “**desarrollo de un modelo matemático para describir el crecimiento de arándanos biloxi con variables ambientales en un sistema cerrado y abierto**”.

Propósito: La intención de este documento es construir los requerimientos funcionales y no funcionales para el proyecto “Desarrollo de un modelo matemático para describir el crecimiento de arándanos Biloxi con variables ambientales en un sistema cerrado y abierto.

Alcance: El propósito de definir estos requisitos es proveer al equipo GISTFA de la Universidad de Cundinamarca una base sólida para facilitar mantenimientos y actualizaciones futuras de software. Esta información servirá como fundamento técnico para el desarrollo de proyectos y la mejora continua.

Ámbito del sistema: El sistema permitirá a los usuarios visualizar el crecimiento del cultivo de arándanos, junto con sus variables y metodologías utilizadas para el análisis. El software almacenará los conjuntos de datos y la información relevante del modelo presentado. Dado que el sistema se desarrollará como un módulo de otro sistema más grande, contará con características adicionales. Por ejemplo, el usuario tendrá la opción de acceder a su información y visualizar otros módulos del sistema general. La interfaz se sustentará mediante una arquitectura REST, utilizando conexiones seguras a través del protocolo HTTPS.

- **Definiciones, Acrónimos y Abreviaturas**

Tabla 4

Definición y Acrónimos

Nombre	Descripción
Análisis de Crecimiento	Análisis descriptivo de variables ambientales respecto a la altura
Modelo matemático	Herramienta matemática que por medio de sus resultados permite la toma de decisiones basada en datos

Variables Ambientales

Son aquellas de las cuales se toman datos por medio de sensores para ser analizadas por medio del modelo matemático.

Nota. Elaboración propia.

Tabla 5

Regencias Estándar de los Requerimientos

Título del Documento	Referencia
Standard IEEE 830 - 2008	IEEE

Nota. Autoría propia.

Descripción general: El módulo de reportes brinda una opción de poder analizar el comportamiento del crecimiento de la planta en relación con las variables ambientales, además de poder tener un análisis de datos por medio de gráficas y tablas de datos.

Esto por medio de un Módulo de reportes que permite mostrar esto al usuario los modelos matemáticos usados y los resultados obtenidos por estos, resaltando características y enfoques que tiene cada uno de estos.

Perspectiva del producto: Este software está relacionado a un prototipo de sistema computacional para gestionar y controlar el cultivo de bayas silvestres en un sistema cerrado de recirculación de agua. Por esto, este módulo cumple un rol fundamental para el correcto desarrollo del sistema asociado, brindando el siguiente módulo:

- **Módulo de Reportes**

En este módulo, los usuarios y administradores podrán generar y visualizar reportes que muestren de forma descriptiva el comportamiento de la planta y las distintas variables consideradas relevantes en el crecimiento de la planta.

Características de los usuarios

Tabla 6

Descripción Rol Super Administrador

Tipo de Usuario	Usuario Privilegiado
Nombre de Usuario	Super Administrador
Descripción	Persona que cuenta con los permisos necesarios para poder ejecutar las operaciones requeridas en el sistema, incluyendo la manipulación y creación de contenido, para garantizar el correcto funcionamiento de este.
Nivel de Educación Mínimo Sugerido	Universitario (Pregrado)
Experiencia Requerida	Personal con conocimiento previo en manejo de tecnologías de .NET Core, React, Python y Matlab con énfasis en consumo de API'S y manejo de bases de datos.

Nota. Elaboración propia.

Tabla 7

Descripción Rol Usuario

Tipo de Usuario	Usuario no privilegiado
Nombre de Usuario	Usuario
Descripción	Persona con registro previo en el sistema para la interacción con el modulo
Nivel de Educación Mínimo Sugerido	No requerida
Experiencia Requerida	Persona con conocimientos básicos en herramientas tecnológicas.

Nota. Autoría propia.

Restricciones y Restricciones de Diseño

- Para la interfaz web se utilizará React.
- Se recomienda utilizar navegadores que admitan HTML5, CSS3 y JavaScript para acceder a la interfaz gráfica del módulo.
- Se utilizará el formato JSON para consumir APIs y realizar manipulaciones en las solicitudes de recursos.
- Para el desarrollo del modelo matemático; construcción y ajuste del modelo y manejo de variables y datasets, se usará algún lenguaje de programación como Python o Matlab.

Suposiciones y Dependencias

Este módulo, junto con los demás recursos del sistema principal compuesto por el software, base de datos e interfaz, pueden instalarse en diferentes equipos de cómputo para lograr un buen manejo de concurrencia.

Requerimientos Futuros

- El módulo permitirá predecir el crecimiento de la planta simulando el crecimiento de esta

Requisitos Específicos

Interfaz de usuario: Interfaz de usuario Diseñada para navegadores web, El dispositivo en donde se vaya a dar manejo al módulo debe contar con la capacidad de soportar un navegador web que cumpla con los requisitos que se han definido anteriormente en el documento.

Interfaz de hardware: No aplica

Interfaz de Software

- **Sistema de instalación:** Se selecciona Windows como sistema operativo debido a su sólido soporte para la tecnología .NET y su interfaz que es fácil de usar.

- **Base de Datos:** Se opta por SQL Server como SGDB relacional, ya que permite por medio de herramientas poder realizar el ORM de manera fácil con .NET y así realizar consultas por medio del lenguaje de consultas LINQ.
- **Tecnología .Net:** Se emplea el entorno de Microsoft Visual Studio para la programación del sistema de información por su conjunto de herramientas que son útiles para el desarrollo.

Funciones

Tabla 8

Requerimiento RF-01.

Identificación del Requerimiento	RF-01
Nombre del Requerimiento	Interfaz Web
Roles	Descripción
Usuario, SuperAdministrador	El sistema permitirá visualizar la interfaz principal del módulo de software, dónde se pondrá la navegación a las diferentes vistas, cómo gráficas, tablas o reportes.

Nota. Autoría propia.

Tabla 9

Requerimiento RF-02

Identificación del Requerimiento	RF-02
Nombre del Requerimiento	Visualización de Datasets
Roles	Descripción
Usuario, SuperAdministrador	El sistema permitirá a los usuarios visualizar los datasets usados en el diseño y elaboración del modelo matemático mediante tablas. Adicionalmente permitirá descargar los datasets en al menos dos formatos: csv y json.

Nota. Autoría propia.

Tabla 10
Requerimiento RF-03

Identificación del Requerimiento	RF-03
Nombre del Requerimiento	Visualización de metodología y técnicas usadas para la elaboración del modelo
Roles	Descripción
Usuario, SuperAdministrador	El sistema permitirá visualizar el procedimiento realizado para construir el modelo matemático, desde el origen de los datos, el tratamiento de estos, la selección de modelos y técnicas, el ajuste de estos y la comparación entre el modelo y los datos obtenidos.

Nota. Autoría propia.

Tabla 11
Requerimiento RF-04.

Identificación del Requerimiento	RF-04
Nombre del Requerimiento	Visualizar modelo matemático
Roles	Descripción
Usuario, SuperAdministrador	El sistema permitirá visualizar el modelo junto a sus características, gráficas y parámetros, logrando observar el comportamiento de las variables a través del tiempo y el coeficiente de correlación entre las variables ambientales y la variable objetivo de crecimiento.

Nota. Autoría propia.

Tabla 12
Requerimiento RF-05

Identificación del Requerimiento	RF-05
Nombre del Requerimiento	Generación reportes
Roles	Descripción
Usuario, SuperAdministrador	El sistema permitirá descargar en formato PDF los reportes generados que contendrán la misma información presentada en el módulo web, pero de forma organizada

y concisa con toda la información pertinente del reporte a generar, algunos de los reportes que se plantean generar inicialmente son: comportamiento de las variables, metodología usada y modelo matemático con sus características.

Nota. Autoría propia.

Requerimientos de rendimiento: Concurrencia de Usuarios, el sistema deberá garantizar al menos el acceso a 5 personas simultáneamente para la consulta de la información en el módulo.

Requisitos no funcionales:

- Concurrencia de Usuarios
- Protección de Datos
- Programación del sistema
- Arquitectura del Sistema
- Diseño mockups
- Desarrollo modelo matemático

Especificación del Diseño

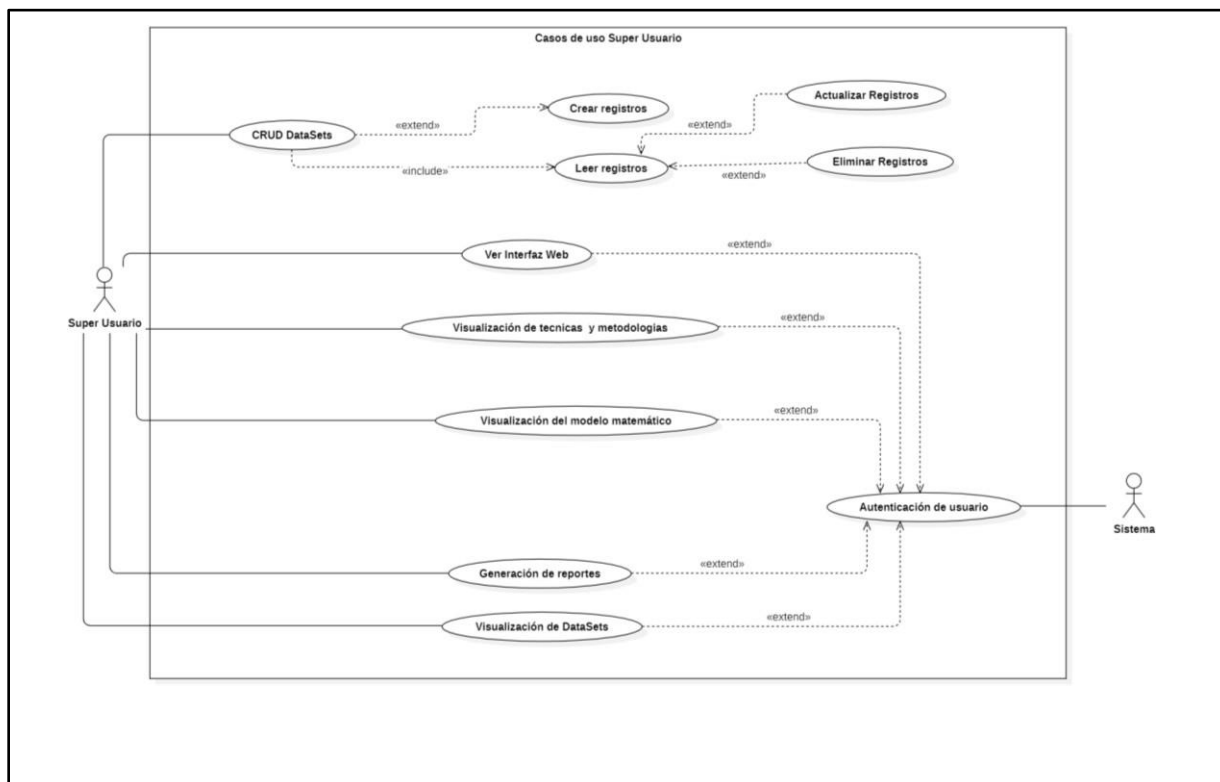
En este apartado se presentan los distintos diagramas elaborados para el proyecto, incluyendo los casos de uso para el superusuario y el usuario, así como los diagramas de secuencia y actividad correspondientes a cada caso de uso. Estos diagramas permiten comprender mejor la interacción del usuario con el sistema y el funcionamiento de las diferentes operaciones, detallando la comunicación entre los servicios y el servidor.

2.2.1. Diagramas de Casos de Uso

Diagrama de Casos de Uso Super Usuario

Figura 7

Diagrama de casos de uso para el rol de Super Usuario.



Nota. Autoría propia.

En el diagrama se cuenta con dos actores el super usuario el cual es aquel usuario que tiene permisos avanzados para poder gestionar y visualizar información del sistema, y el sistema el cual cumple el rol de proporcionar la autenticación y gestión de datos.

a. Descripción.

Tabla 13

Tabla descriptiva de los casos de uso para el rol de Super Administrador.

Componente del Diagrama	Descripción
Relaciones	Se emplean relaciones tipo extend para señalar casos de uso que se complementan con otros. También se usan include en la parte de lectura de registros, describiendo que un paso obligatorio dentro del CRUD.
Casos de Uso	
CRUD Datasets	Este caso de uso permite realizar operaciones de crear, leer, actualizar y eliminar el registro de registros en el conjunto de datos, además de esto se divide en 4 casos de uso con diferentes relaciones esto con el fin de poder tener un mejor entendimiento de cómo funciona este caso de uso.
Crear registro (Extend)	Permite al super usuario crear nuevos registros en el sistema.
Leer registro (Include)	Acción necesaria dentro del CRUD para realizar consultas de los datasets almacenados
Actualizar registro (Extend)	Permite realizar modificaciones a los registros existentes.
Eliminar registro (Extend)	Permite eliminar registros
Ver interfaz Web (Extend)	Permite al usuario tener acceso a al interfaz grafica del módulo del sistema.
Visualización de técnicas y metodologías (Extend)	Permite al usuario ver el apartado en donde se muestran las técnicas y metodologías empleadas en la construcción y estructuración del modelo matemático

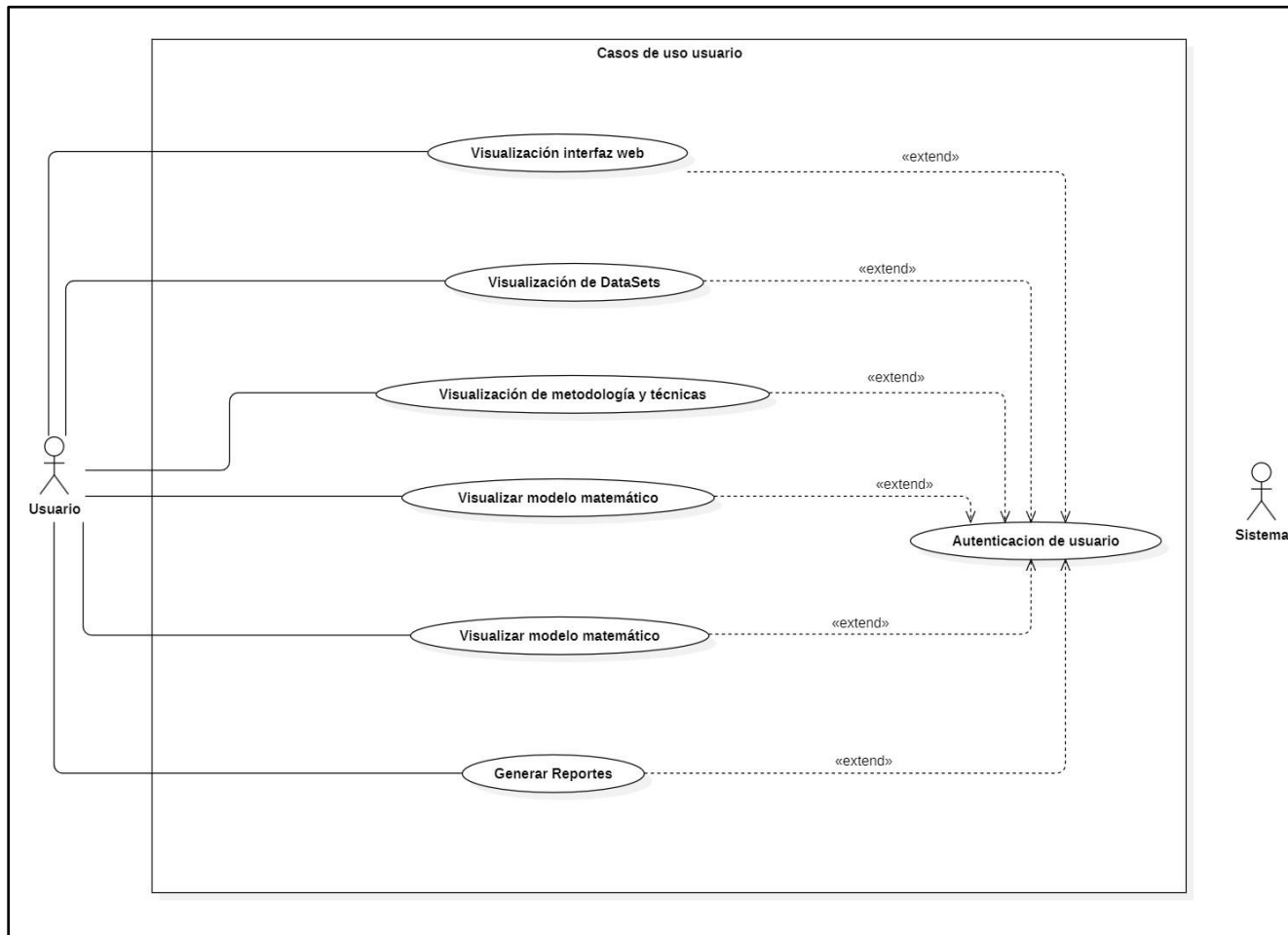
Visualización del modelo matemático (Extend)	Permite visualizar el modelo matemático que se ha usado para la descripción del crecimiento de las plantas
Visualización de técnicas de análisis de imágenes (Extend)	Apartado en donde se ven herramientas y técnicas de análisis de imágenes que permiten describir y comprender el crecimiento de las plantas
Gestión de reportes (Extend)	Opción para generar reportes con base a los datasets almacenados determinando parámetros que se le solicitan al usuario
Visualización de Datasets (Extend)	Muestra el conjunto de datos que se han ido tomando
Autenticación de usuario	Es un caso de uso fundamental ya que varios procesos requieren de una autenticación previa para poderse ejecutar

Nota. Elaboración propia.

Diagrama de Casos de Uso Usuario

Figura 8

Diagrama de casos de uso para el rol de Usuario.



Nota. Elaboración propia

a. Descripción.

Tabla 14

Tabla descriptiva de los casos de uso para el rol de Usuario.

Componente del Diagrama	Descripción
Relaciones	Se emplean relaciones tipo extend para señalar casos de uso que se complementan con otros. También se usan include en la parte de lectura de registros, describiendo que un paso obligatorio dentro del CRUD.
Actores	<p>Usuario: Es aquel actor que interactúa con el sistema para acceder a información, visualizar modelos matemáticos y generar reportes.</p> <p>Sistema: Plataforma encargada de procesar solicitudes, autenticar usuarios y gestionar datos</p>
Casos de Uso	
Visualización de interfaz web (Extend)	Permite al usuario acceder a la interfaz web del sistema
Visualización de DataSets (Extend)	Permite inspeccionar conjunto de datos almacenados en el sistema
Visualización del modelo matemático (Extend)	Permite acceder a la representación e información del modelo matemático
Generación de reportes (Extend)	Funcionalidad para generar reportes basados en los datasets del sistema
Autenticación de usuario	Caso de uso clave del sistema que extiende a otras acciones que requieren validación de identidad

Nota. Autoría propia.

2.2.2. Diagramas de Secuencia

Componentes de diagrama de secuencia

Tabla 15

Descripción de los componentes de un diagrama de secuencia

Componente	Función del componente
Línea de vida Usuario	Representa al actor que interactúa con el sistema, siendo el punto de inicio del flujo de la comunicación
Línea de vida Interfaz	Actúa como el intermediario entre el usuario y el sistema, recibiendo solicitudes y enviándola a componentes internos
Línea de vida Infraestructura	Medio de comunicación entre la interfaz y la aplicación.
Línea de vida Aplicación	Contiene la lógica del negocio y se encarga de procesar las solicitudes recibidas desde la interfaz a través de la infraestructura.
Línea de vida Dominio	Es consultado por la aplicación para recuperar, modificar o validar datos sean las necesidades del proceso.
Request	Representa la solicitud inicial enviada por el usuario para realizar una acción específica en el sistema.
Fetch Data	Solicitud que se ejecuta dentro del sistema cuando la aplicación necesita recuperar datos.
Request response	Es la respuesta generada por el sistema tras procesar la solicitud. Dependiendo de la

solicitud, esta respuesta puede incluir la información solicitada, mensaje de error o confirmación de acciones realizadas.

QueryDB

Representa la consulta directa a la base de datos que se ejecuta cuando la aplicación necesita acceder a información almacenada.

Punto Final

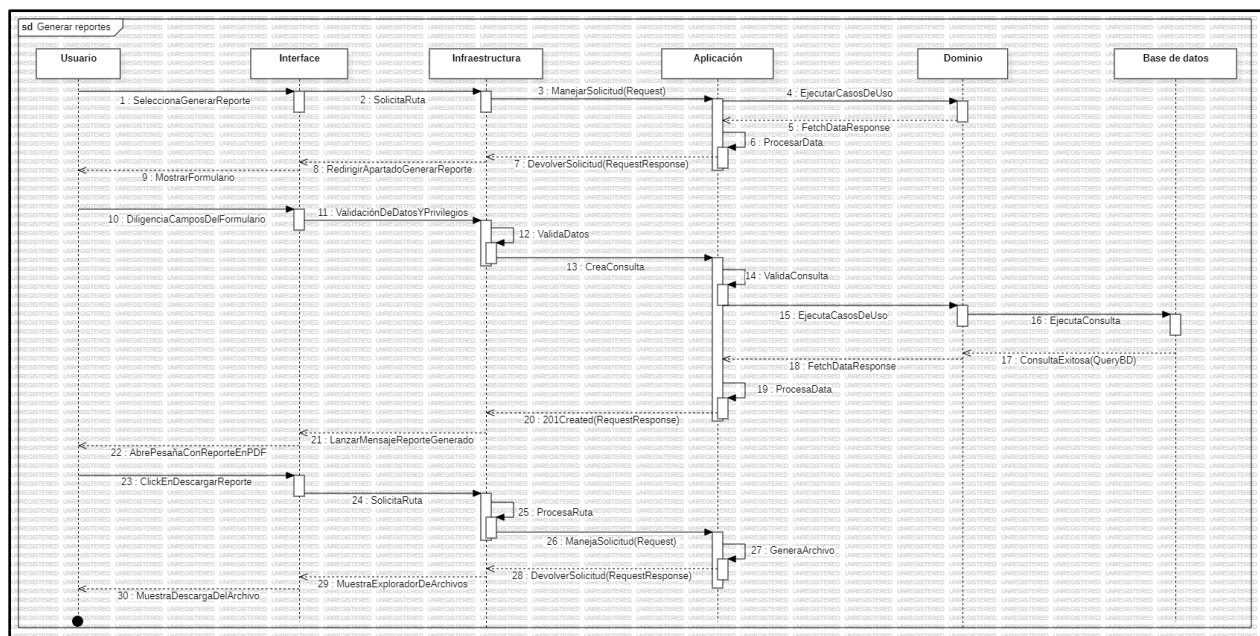
Representa el final del flujo de secuencia del proceso solicitado por el usuario.

Nota. Autoría propia.

En la Tabla 15 se describen las líneas de vida presentes en el diagrama de secuencia, junto con los tipos de solicitudes que ejecuta cada una para interactuar entre sí mediante el envío y recepción de datos. Además, se detalla la estructura de la arquitectura utilizada en el proyecto compuesta por las diferentes ejecuciones detalladas en los diagramas.

Diagrama de secuencia Generar reportes

Figura 9
Diagrama de secuencia Generar reportes

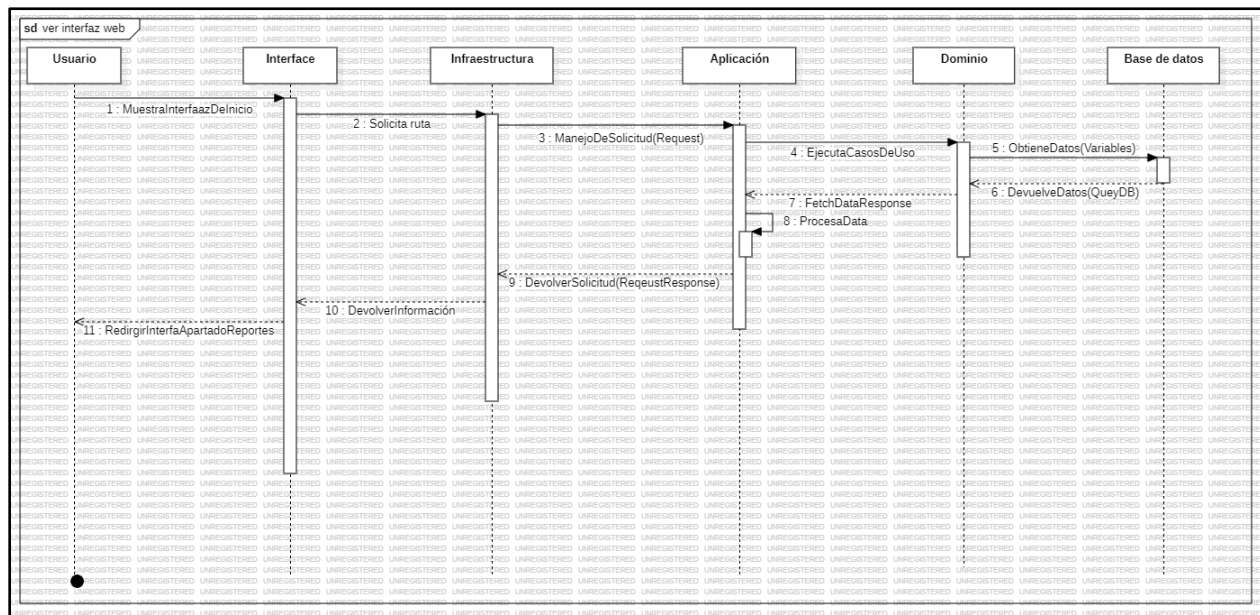


Nota. Autoría propia.

En la Figura 9 se muestra el diagrama de secuencia, donde cada línea de vida del proyecto interactúa mediante consultas. El proceso comienza en la línea del usuario, con el servidor enviando respuestas 200 OK. Para la generación del reporte, el sistema ejecuta el caso de uso y la consulta a la base de datos (QueryDb), recuperando los datos necesarios, generando el archivo y permitiendo su descarga. Finalmente, se muestra al usuario un mensaje de confirmación de la descarga.

Diagrama de secuencia: Ver interfaz

Figura 10
Diagrama de secuencia para ver interfaz.

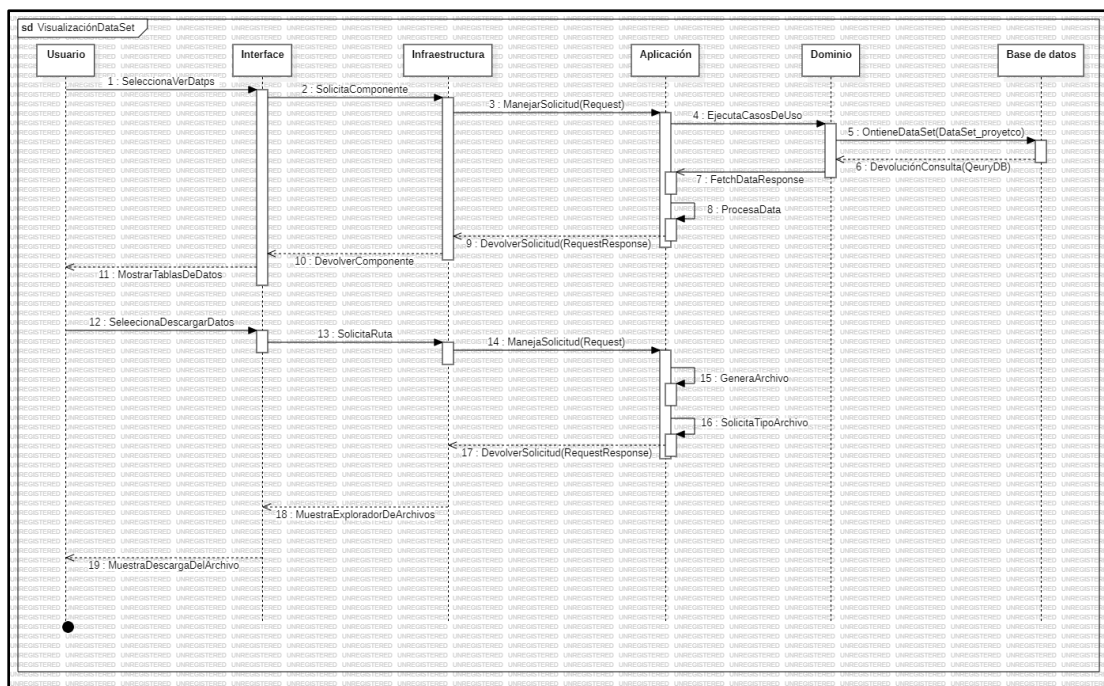


Nota. Elaboración propia.

En la Figura 10 se muestra la secuencia del caso de uso "Ver interfaz web", donde el usuario solicita la ruta correspondiente. A lo largo del proceso, cada línea de vida ejecuta solicitudes y consultas para cargar y mostrar el módulo de reportes al usuario.

Diagrama de secuencia: Visualización DataSet

Figura 11
Diagrama de secuencia para visualización dataset.

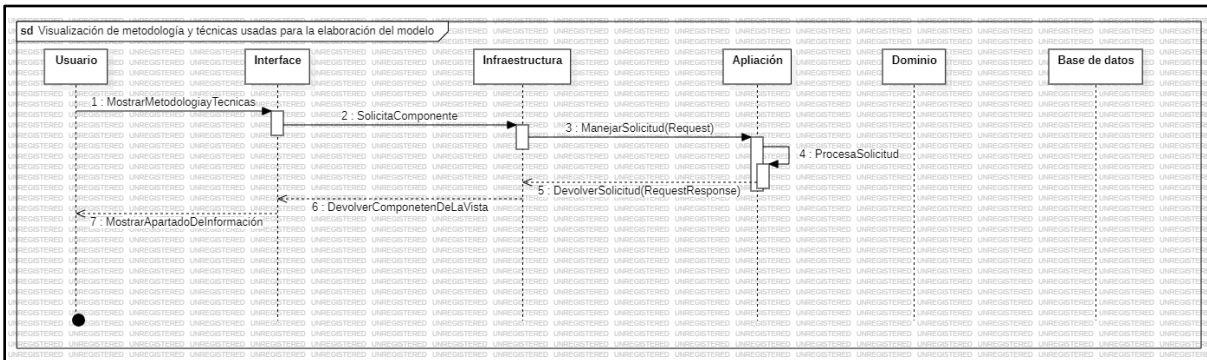


Nota. Autoría propia.

En la Figura 11 se muestra la secuencia de interacción entre las diferentes líneas de vida en la tarea "Visualización de datasets". Durante el proceso, se ejecutan validaciones y consultas a través de cada línea de vida. Además, se representa la secuencia de descarga del dataset cuando el usuario realiza esta acción.

Diagrama de secuencia: Técnicas Metodologías y técnicas usadas para la elaboración del modelo

Figura 12
Diagrama de secuencia: Técnicas Metodologías y técnicas usadas para la elaboración del modelo.

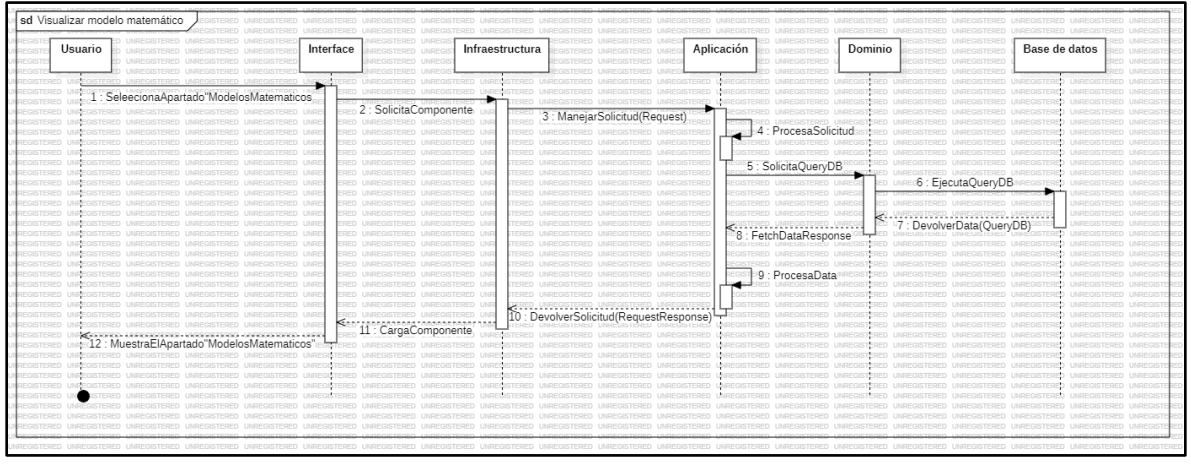


Nota. Elaboración propia.

En la Figura 12 se muestra la secuencia de carga del componente "Técnicas metodologías y técnicas usadas para la elaboración del modelo". Durante el proceso, cada línea de vida interviene para garantizar la correcta ejecución del cargue hasta llegar a la línea de aplicación regresando a la línea de usuario hasta mostrar el componente dentro de la vista.

Diagrama de secuencia Visualización modelo matemático

Figura 13
Diagrama de secuencia para visualización modelo matemático

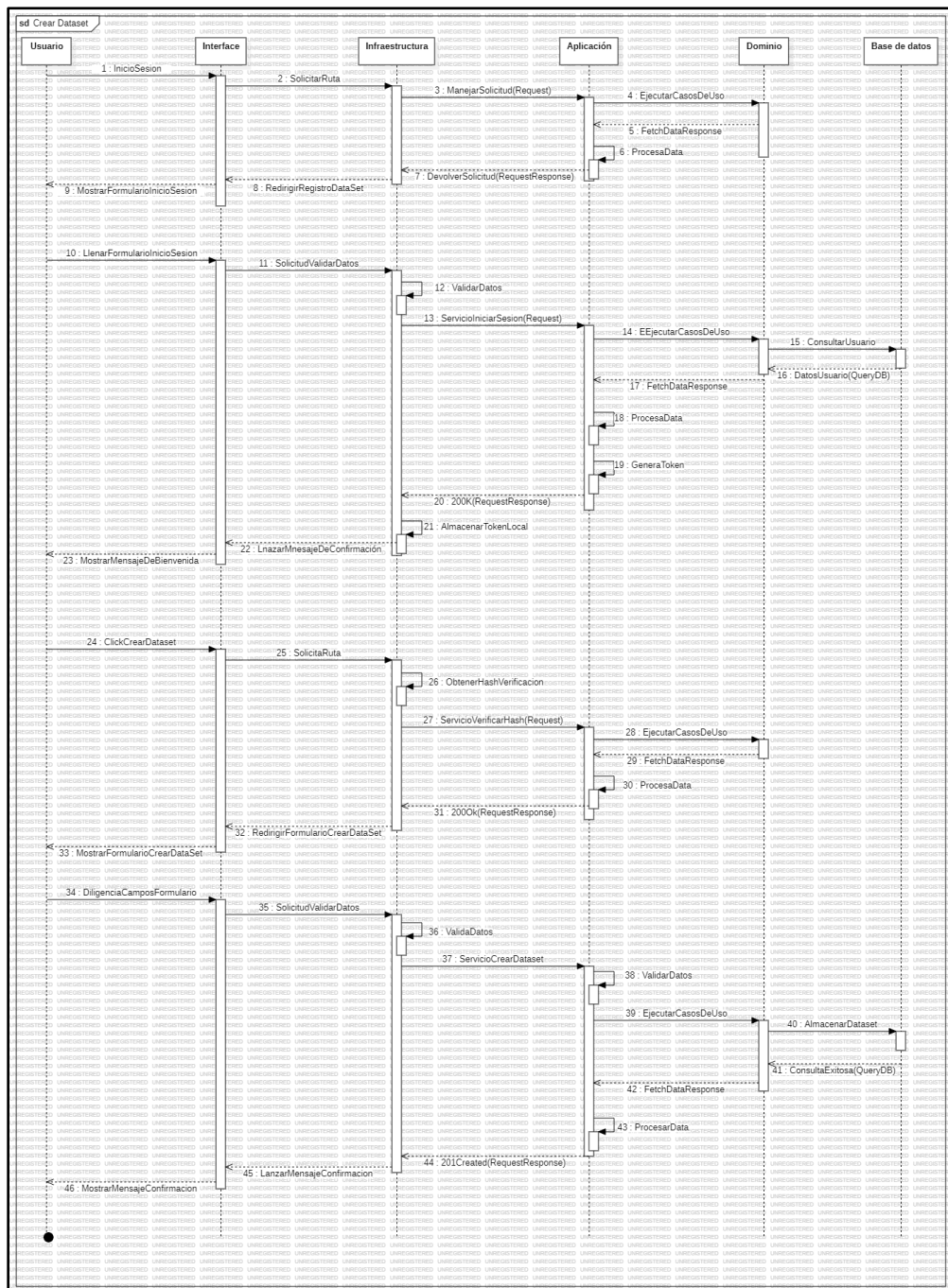


Nota. Elaboración propia.

En la Figura 13 se representa el flujo de validaciones y consultas que el sistema ejecuta cuando el usuario selecciona el componente "Modelo Matemático". Durante este proceso, el sistema verifica los permisos y realiza las consultas necesarias para recuperar la información correspondiente. A continuación, se muestran los modelos matemáticos disponibles y se procede a la carga de los datos extraídos desde la línea de base de datos. Este flujo garantiza que el usuario acceda a la información de manera estructurada y más comprensible para su análisis.

Diagrama de secuencia Crear Dataset

Figura 14
Diagrama de secuencia para Crear Dataset

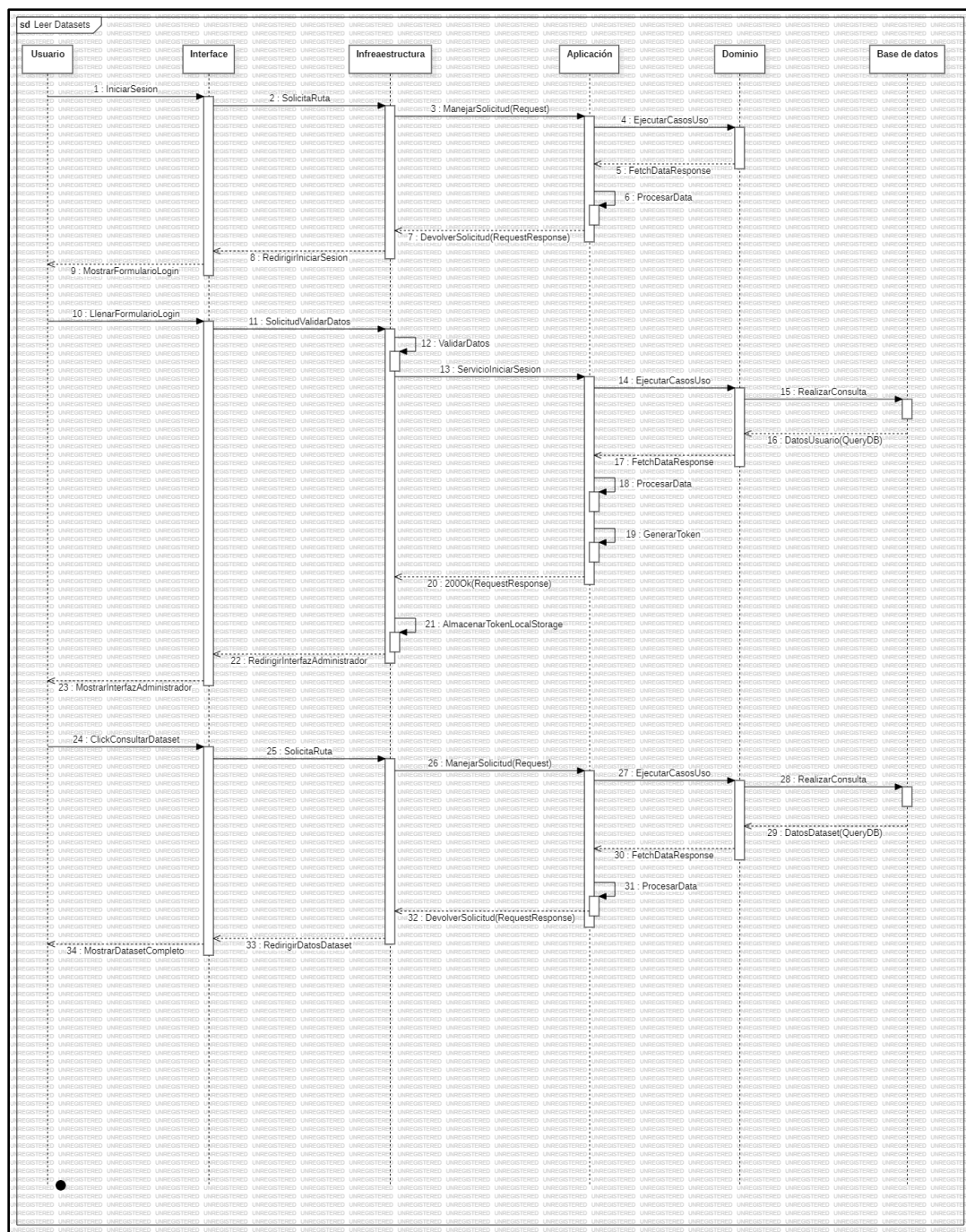


Nota. Elaboración propia.

En la Figura 14 se muestra la secuencia de actividades que el usuario lleva a cabo al interactuar con el sistema para crear nuevos datasets. Durante este proceso, cada línea de vida ejecuta solicitudes y respuestas, permitiendo la validación de la información ingresada. El sistema realiza autenticaciones y validaciones de identidad mediante actividades como la tokenización y la ejecución de consultas a la base de datos (QuerysDb). Además, se garantiza que las respuestas proporcionadas permitan al usuario continuar con el proceso de creación del dataset de manera segura y estructurada.

Diagrama de secuencia: Leer Dataset

Figura 15
Diagrama de secuencia para Leer Dataset

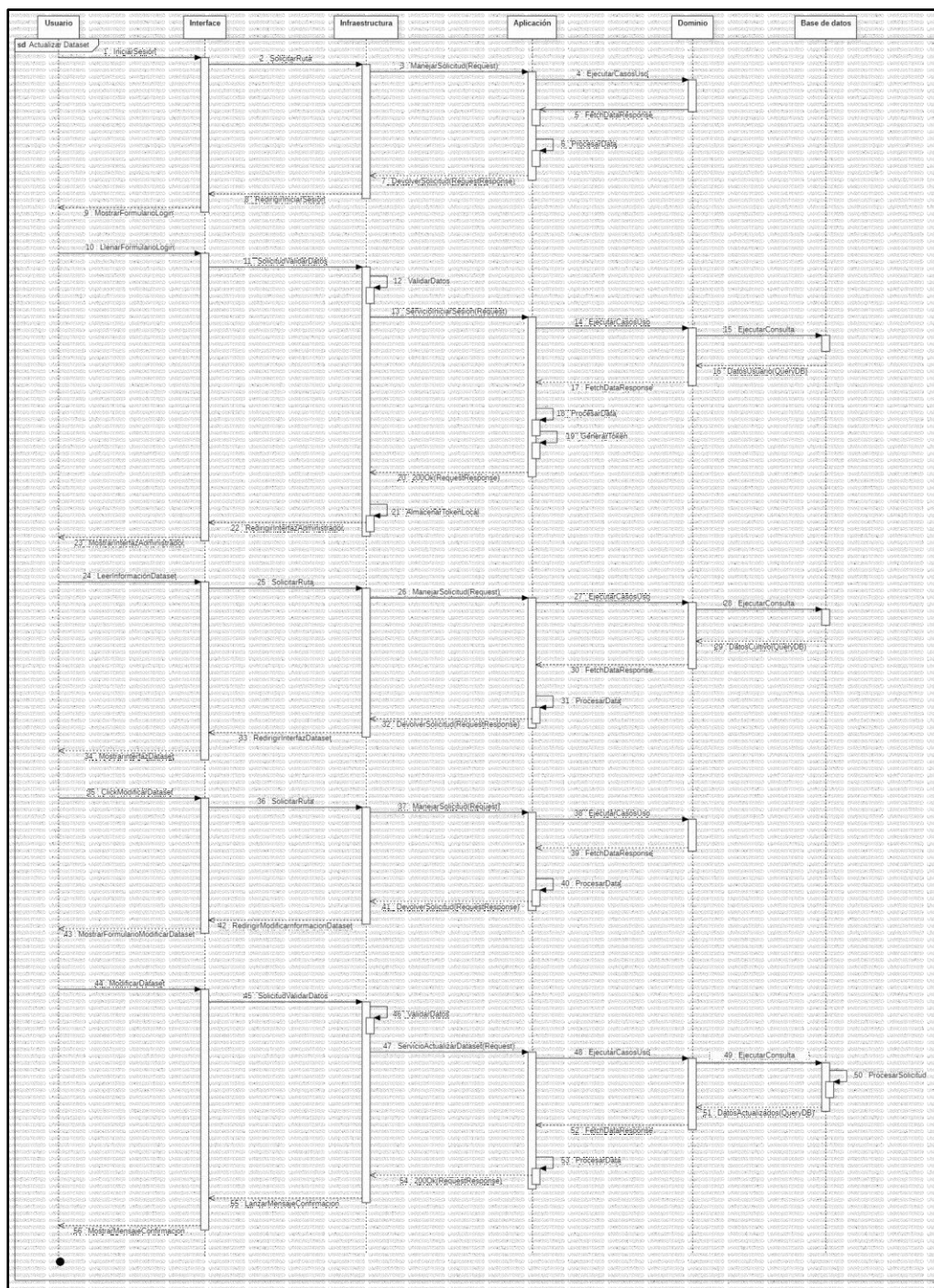


Nota. Autoría propia.

En la Figura 15 se muestra la secuencia de actividades que el usuario lleva a cabo al interactuar con el sistema para realizar la búsqueda de un dataset. El proceso comienza con la autenticación del usuario, seguida de una serie de validaciones y consultas que se ejecutan a lo largo de las diferentes líneas de vida. El sistema verifica los permisos y realiza las consultas necesarias a la base de datos (QuerysDb) para recuperar y mostrar la información solicitada de manera estructurada y accesible para el usuario.

Diagrama de secuencia: Actualizar Dataset

Figura 16
Diagrama de secuencia para Actualizar Dataset.

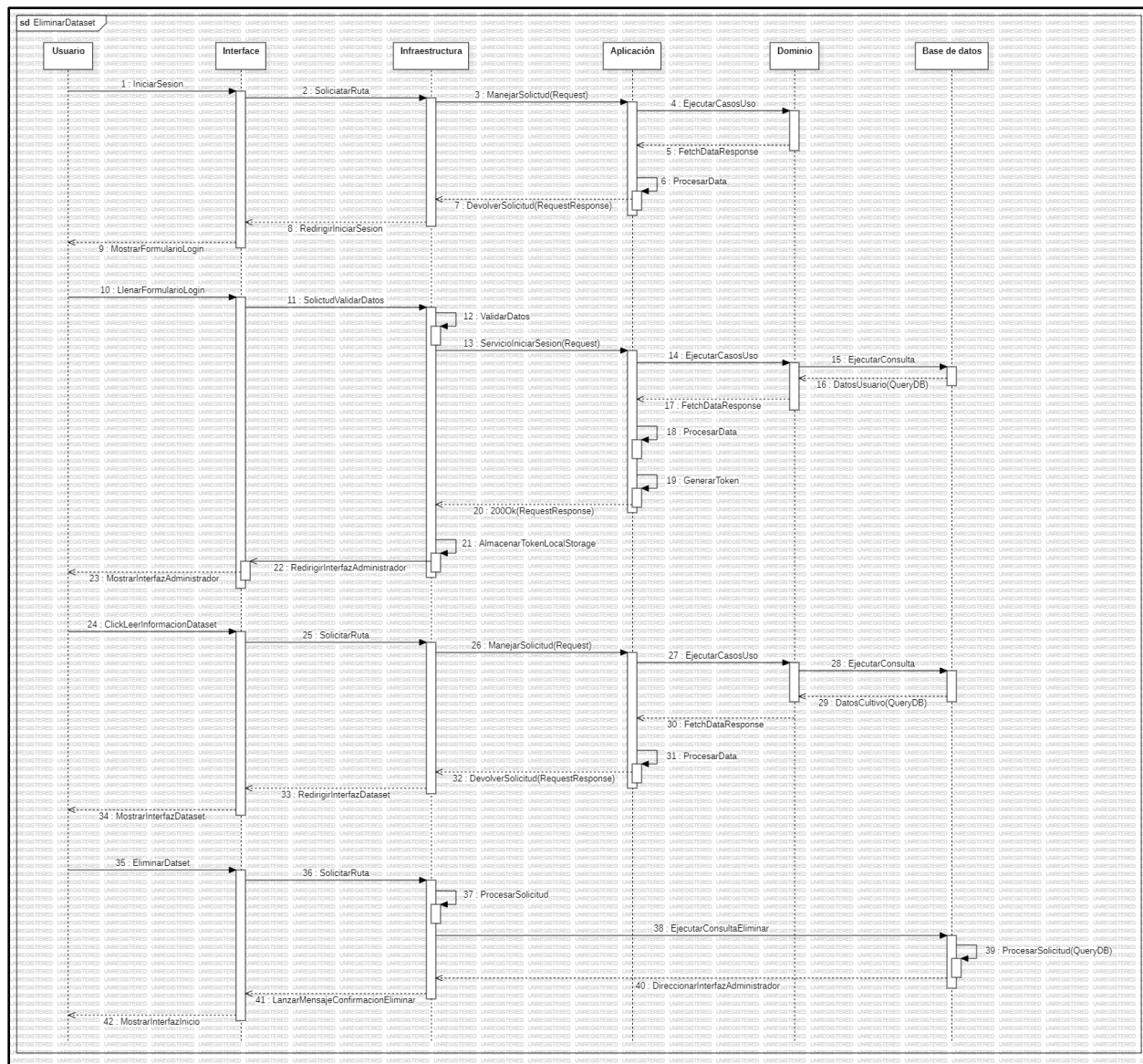


Nota. Elaboración propia.

En la Figura 16 se muestra la secuencia de acciones que el usuario lleva a cabo al interactuar con el sistema para actualizar un dataset. El proceso inicia con la autenticación del usuario y la validación de sus permisos. A continuación, el sistema ejecuta una serie de consultas y verificaciones para garantizar la integridad de los datos. Durante la actualización, el usuario recibe respuestas que le permiten seguir el estado del proceso. Finalmente, una vez completada la actualización, el sistema genera un mensaje de confirmación indicando que la operación se ha realizado con éxito.

Diagrama de secuencia Eliminar Dataset

Figura 17
Diagrama de secuencia Eliminar Dataset



Nota. Autoría propia.

En la Figura 17 se muestra el proceso final del CRUD de datasets: la eliminación de un dataset. El flujo inicia con la autenticación del usuario, asegurando que tenga los permisos necesarios para realizar esta acción. A continuación, el sistema ejecuta una serie de validaciones

y consultas para verificar la existencia del dataset y confirmar que su eliminación no afectará otras operaciones. Durante el proceso, el usuario recibe retroalimentación sobre el estado de la solicitud. Finalmente, el sistema confirma la eliminación exitosa mediante un mensaje de confirmación, asegurando que la operación se ha completado correctamente.

2.2.3. Diagramas de Actividades

Componentes de los diagramas de actividad.

Tabla 16

Descripción de los componentes

Componente	Función del componente
Carril Usuario	Este carril encierra las actividades que realiza el rol de usuario
Carril Sistema	Este carril encierra las actividades que ejecuta el sistema
Punto inicial	Marca el inicio del flujo de acciones
Cuadro de decisión	Indica el flujo de decisión, dependiendo la validación de los datos ejecuta una acción
Acción	Espacio donde se describe la acción que se ejecuta
Control de flujo	Señalización del sentido en que se ejecuta la acción
Final	Señala el final del flujo de acciones

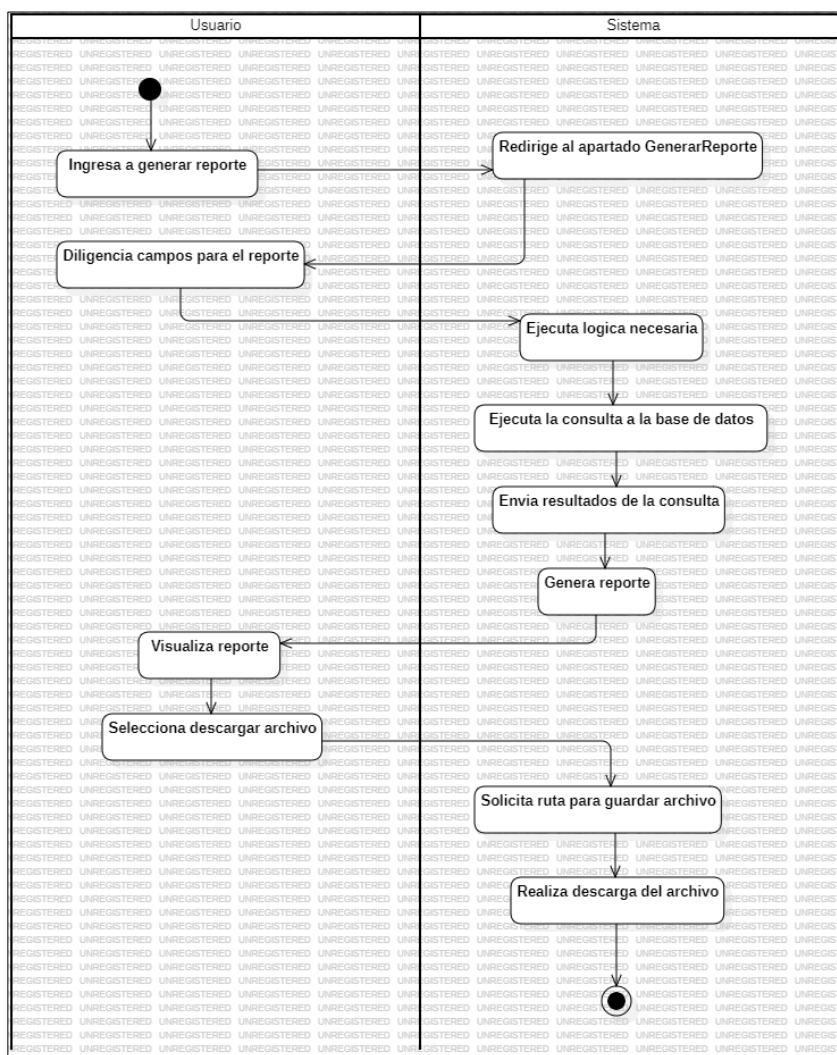
Nota. Elaboración propia.

En la tabla 16 se presenta los componentes del diagrama de actividad y una pequeña descripción de lo que representa cada uno de estos siguiendo en flujo de las actividades.

Diagrama de actividad: generar reportes

Figura 18

Diagrama de actividad para generar reportes.



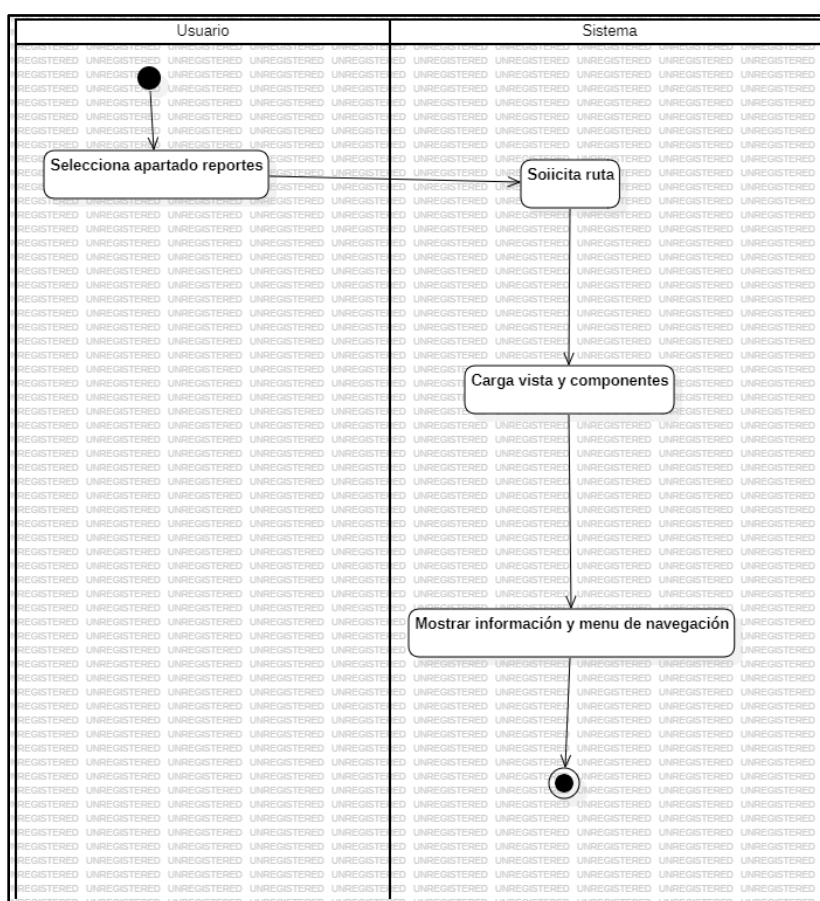
Nota. Autoría propia.

En la figura 18 se muestra el diagrama de actividad para la generación de reportes, detallando cada paso desde el inicio hasta la finalización del proceso. Además, se ilustran las actividades que el sistema ejecuta en respuesta a las acciones solicitadas por el usuario para poder generar un reporte de los datos de los cultivos.

Diagrama de actividad: ver interfaz web

Figura 19

Diagrama de actividad para ver interfaz web.



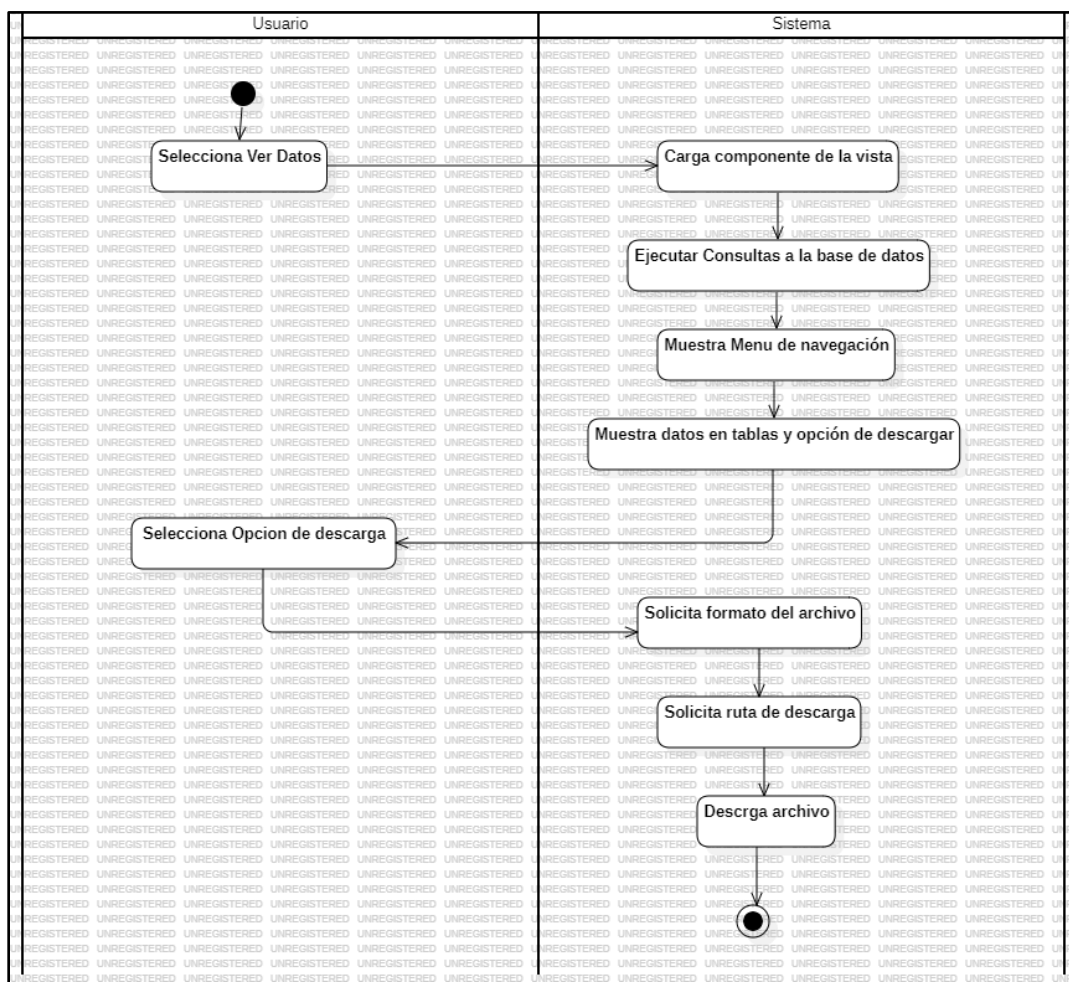
Nota. Elaboración propia.

En la figura 19 se muestra el diagrama de actividad para la acción de visualizar la interfaz web. Este diagrama representa la secuencia de actividades iniciadas por el usuario y ejecutadas por el sistema en respuesta a sus solicitudes cargando los componentes e información de la página.

Diagrama de actividad: visualización de dataset

Figura 20

Diagrama de actividad para visualizar dataset



Nota. Elaboración propia.

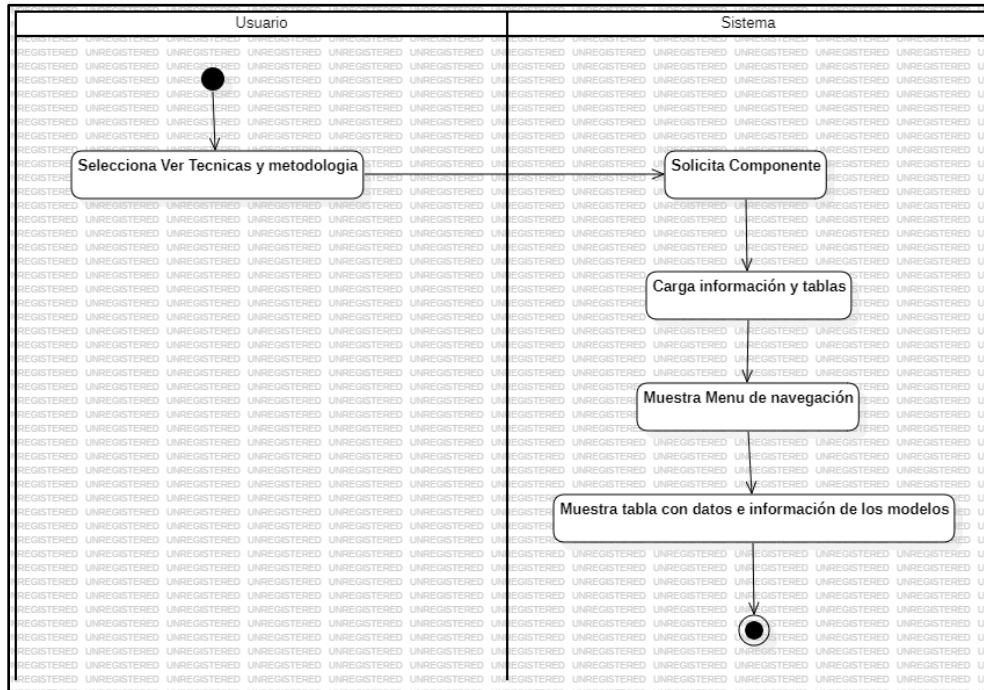
En la figura 20 se muestra la interacción del usuario con el sistema para la visualización de datasets. Además, se incluye la opción de descargar los datasets mostrados

en la interfaz. El proceso finaliza cuando el sistema entrega el archivo descargado con la información solicitada por el usuario.

Diagrama de actividad: visualización metodologías y técnicas

Figura 21

Diagrama de actividad para la visualización metodologías y técnicas.



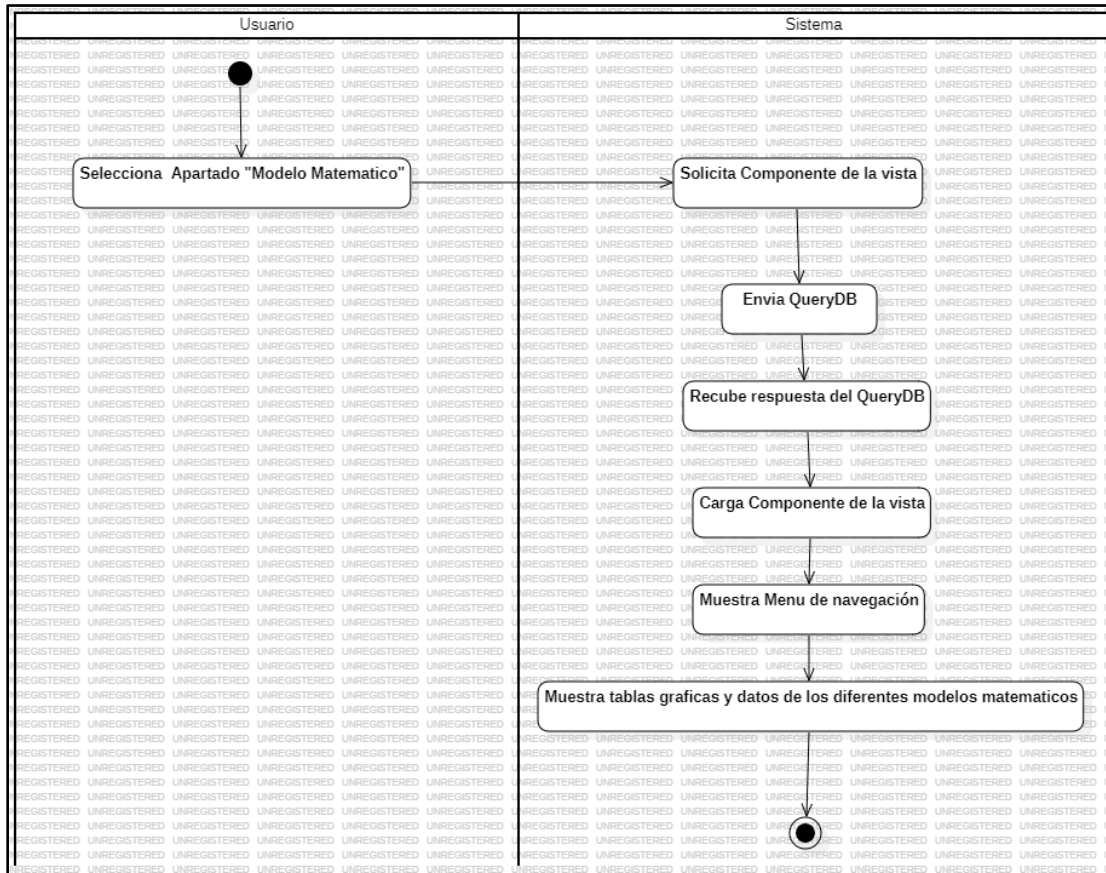
Nota. Elaboración propia.

El proceso de visualización de las metodologías y técnicas empleadas en el modelo matemático se describe a través de una serie de actividades, como se muestra en la Figura 21. En este flujo, el usuario interactúa con el sistema solicitando la carga de la vista, tras lo cual el sistema procesa y muestra los componentes correspondientes. El proceso finaliza cuando la información solicitada se presenta en la interfaz.

Diagrama de actividad: visualización modelo matemático

Figura 22

Diagrama de actividad para la visualización del modelo matemático

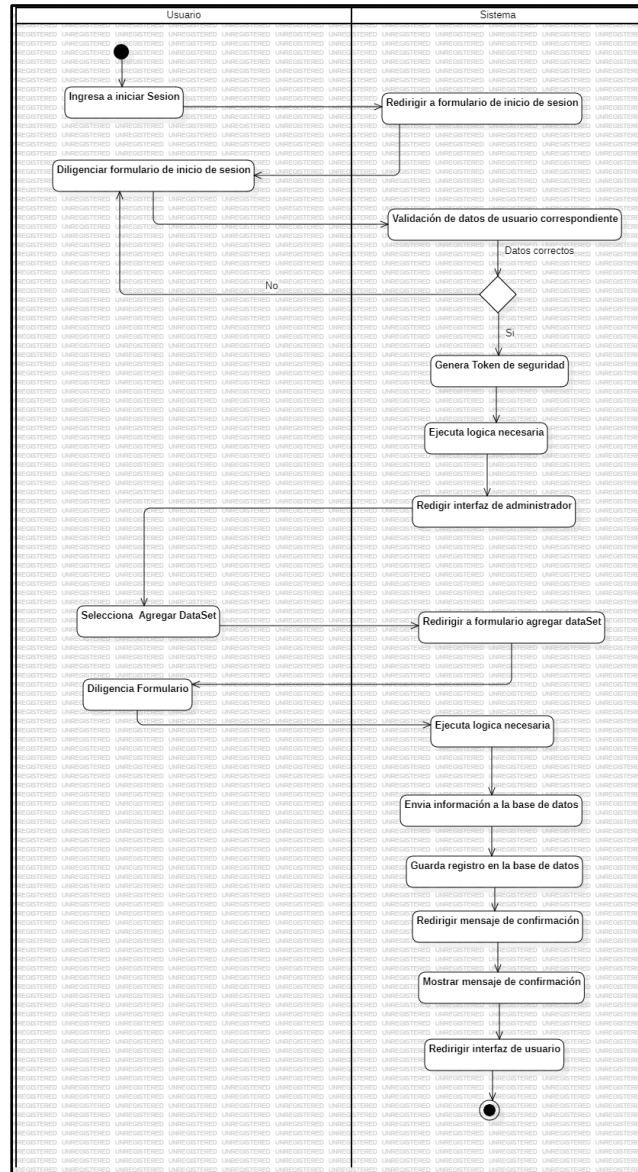


Nota. Autoría propia.

El flujo de actividades para mostrar el modelo matemático empleado para poder describir el crecimiento de la planta se muestra en la figura 22, en donde el usuario selecciona el apartado en donde se muestra el modelo matemático y ahí el sistema realiza el cargue de los datos ejecutando lo necesario para poder mostrar la información de manera dinámica para el usuario.

Diagrama de actividad: Crear Dataset

Figura 23
Diagrama de actividad para crear el dataset

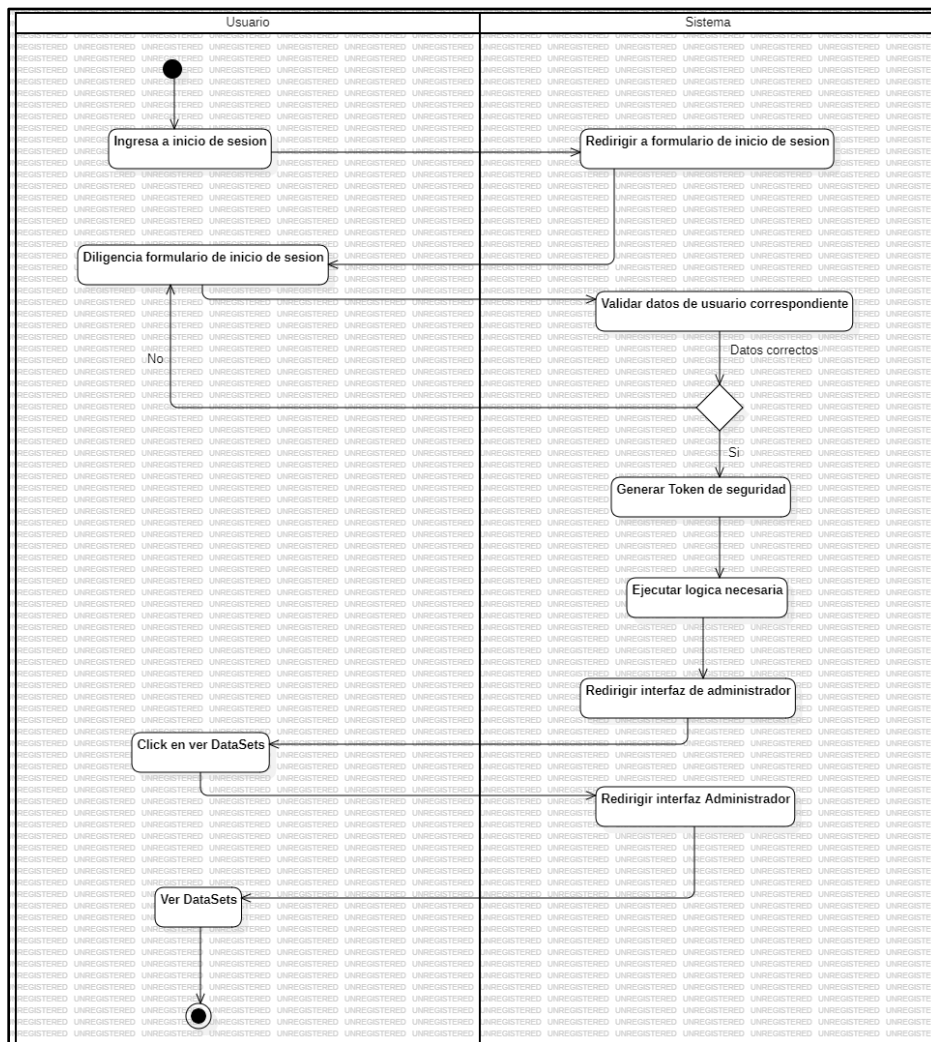


Nota. Autoría propia.

En la Figura 23 se muestran las actividades que el usuario lleva a cabo al interactuar con el sistema para crear nuevos datasets. Durante este proceso, el sistema realiza una serie de validaciones y proporciona respuestas que permiten al usuario continuar. Además, se implementan mecanismos de autenticación y validación de identidad mediante actividades como la tokenización y la ejecución de consultas a la base de datos (QuerysDb).

Diagrama de actividad: Leer Dataset

Figura 24
Diagrama de actividad para leer el dataset



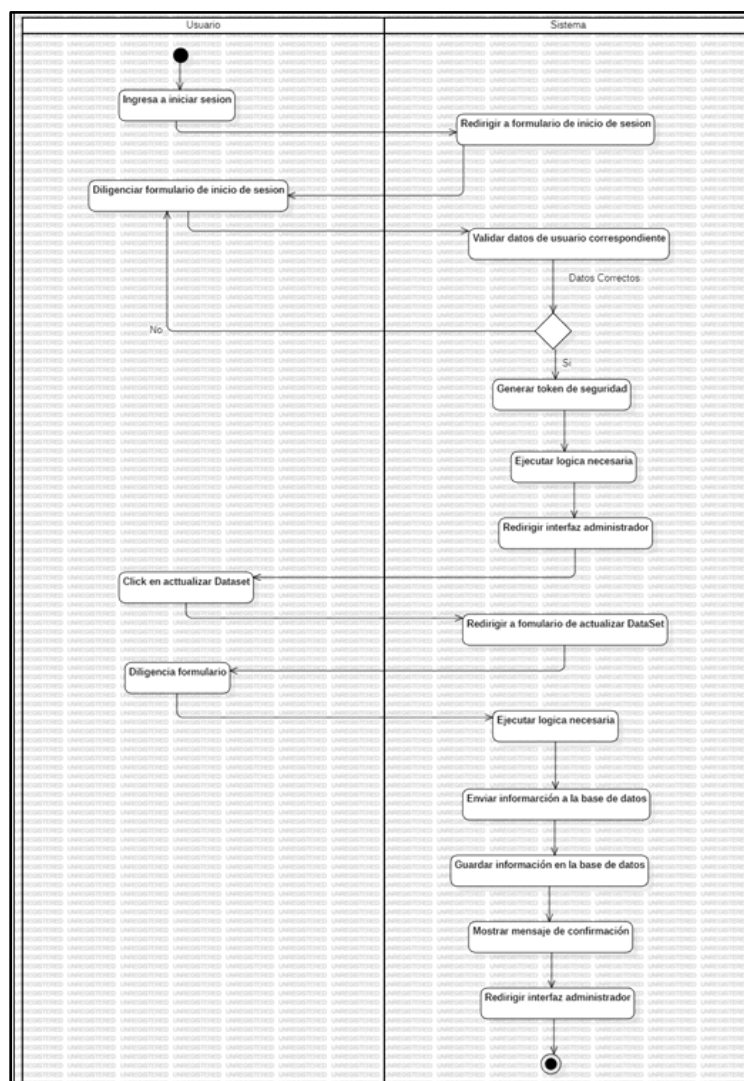
Nota. Autoría propia.

En la Figura 24 se muestra la secuencia de actividades que el usuario realiza al interactuar con el sistema para buscar un dataset. El proceso incluye la autenticación del usuario, junto con las validaciones y consultas que el sistema debe ejecutar para permitir la visualización de la información solicitada.

Diagrama de actividad: Actualizar dataset

Figura 25

Diagrama de actividad para Actualizar dataset

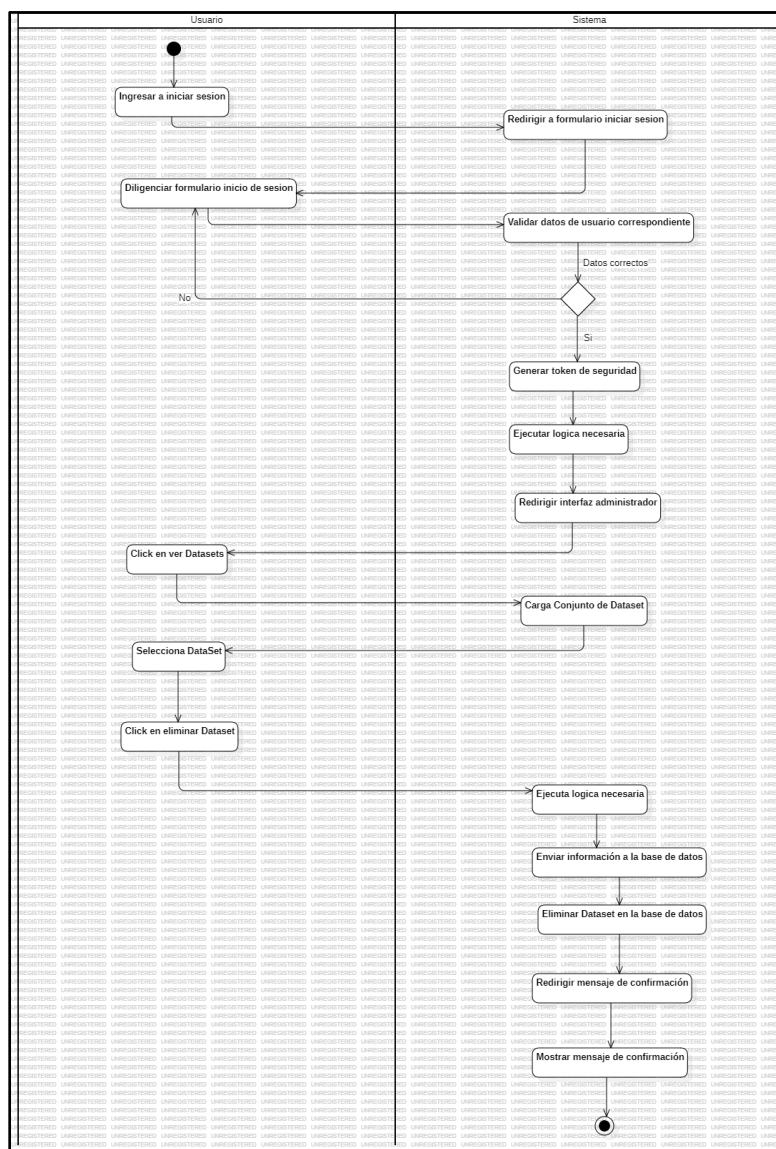


Nota. Autoría propia.

En la Figura 25 se muestra el conjunto de acciones que el usuario ejecuta al interactuar con el sistema para actualizar un dataset. Durante el proceso, el sistema realiza una serie de validaciones y consultas, proporcionando respuestas visibles para el usuario. Finalmente, la actualización concluye con un mensaje de confirmación.

Diagrama de actividad Eliminar Dataset

Figura 26
Diagrama de actividad para eliminar el dataset.



Nota. Elaboración propia.

En la Figura 26 se muestra el último proceso del CRUD de datasets, la eliminación de un dataset. El flujo comienza con la autenticación del usuario en el sistema, seguido de una serie de validaciones y consultas ejecutadas por el sistema para procesar la solicitud. Finalmente, el proceso concluye con un mensaje de confirmación indicando que la eliminación se ha realizado correctamente.

2.3. Diseños de Casos de Prueba

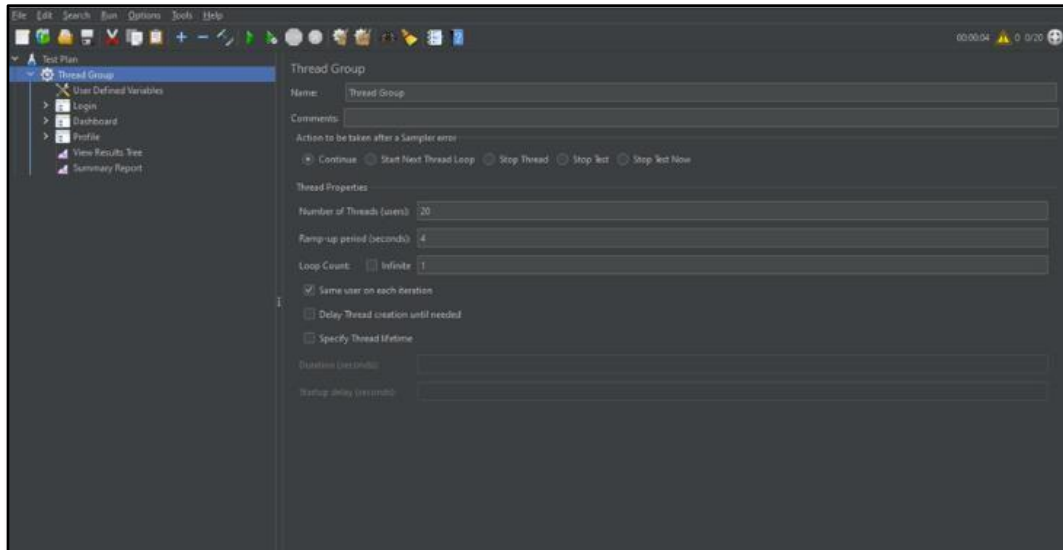
Para el desarrollo de las pruebas se realizaron cargue de los datos obtenidos, mostrando las gráficas, detalles e información relacionada con el modelo, en donde el usuario pueda visualizar el modelo matemático y demás apartados, fundamentalmente la generación de reportes.

Se validó la integración y compatibilidad con el software principal ya que, al ser un módulo de este, se validó que no genera conflictos y que el cargue de información se hiciera de manera correcta, permitiendo la visualización de graficas, información, datasets y descargue de reportes que es lo fundamental del módulo a implementar.

2.4. Pruebas de Estrés y Carga (JMETER)

El presente informe tiene como objetivo documentar los resultados obtenidos durante las pruebas de carga realizadas utilizando Apache JMeter. Estas pruebas fueron diseñadas para evaluar el desempeño y la capacidad de respuesta de la aplicación bajo un escenario de concurrencia, simulando múltiples usuarios accediendo simultáneamente a los servicios del sistema.

Figura 27
Configuración Thread Group (20 hilos 4 segundos)



Nota. Elaboración propia.

Se configuró un grupo de hilos (usuarios) con un total de 20 hilos y un periodo de rampa de 4 segundos, lo cual permitió distribuir la carga de forma progresiva. Las pruebas se enfocaron en los distintos puntos críticos del sistema, incluyendo módulos de autenticación, visualización de dashboard, consulta de información, y actualización de datos. El objetivo fue determinar la estabilidad, el tiempo de respuesta y la capacidad del sistema para manejar múltiples solicitudes simultáneas sin degradación del servicio.

Figura 28
Resultados Obtenidos

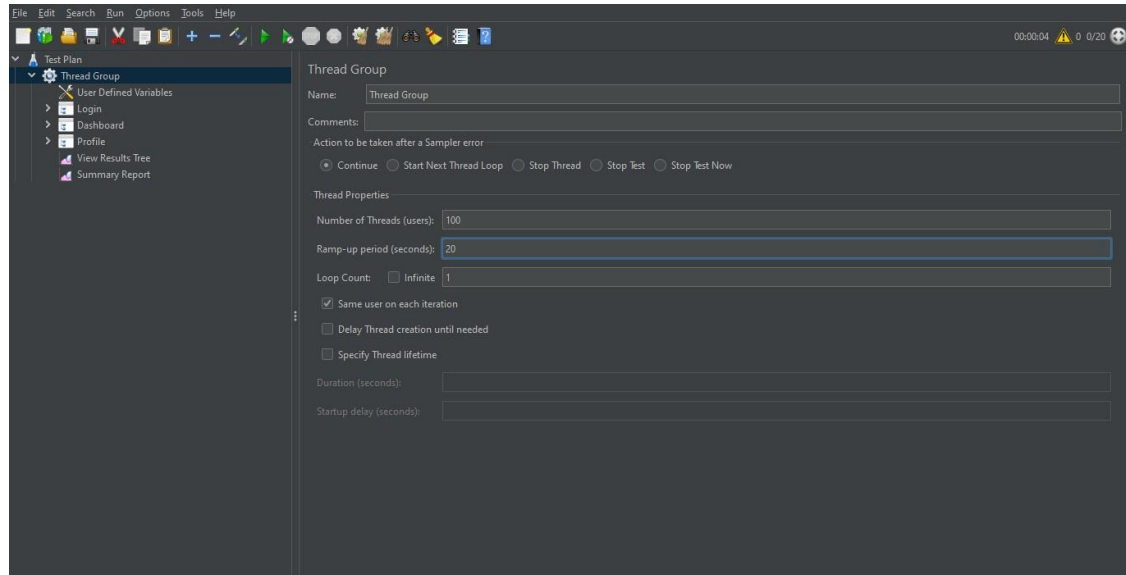
Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
/api/v1/User/login	20	610	480	1651	254.23	0.00%	4.6/sec	3.87	1.23	855.6
Login	20	610	480	1651	254.23	0.00%	4.6/sec	3.86	1.23	855.6
/api/v1/Crop/list	20	97	83	199	5.40	0.00%	5.6/sec	30.18	2.82	5474.0
/api/v1/CultivationSystem/pagination	20	105	96	121	6.24	0.00%	5.6/sec	7.70	3.01	1403.0
/api/v1/Crop/pagination/search/prueba	20	102	92	120	6.43	0.00%	5.6/sec	1.31	3.02	238.0
Dashboard	20	305	272	323	10.55	0.00%	5.6/sec	37.02	6.46	7115.0
/api/v1/User/update	20	115	108	129	5.27	0.00%	5.6/sec	4.69	7.71	855.0
Profile	20	115	108	129	5.27	0.00%	5.6/sec	4.69	7.70	855.0
TOTAL	160	257	85	1651	246.65	0.00%	33.7/sec	72.64	27.22	2206.2

Nota. Elaboración propia.

Durante la ejecución de las pruebas de carga, se realizaron un total de 160 solicitudes distribuidas en diferentes endpoints del sistema. Los resultados obtenidos indican un tiempo de respuesta promedio de 257 milisegundos, sin presentarse errores en la ejecución de las solicitudes (0% de error). La prueba más exigente en cuanto a tiempo de respuesta fue el endpoint `/api/V1/User/login`, con un máximo de 1651 ms y un promedio de 610 ms.

El sistema mantuvo una tasa de transferencia estable con un throughput promedio de 33.1 solicitudes por segundo, y no se evidenciaron caídas ni cuellos de botella que comprometieran la estabilidad general. En resumen, el comportamiento del sistema bajo esta carga específica fue óptimo y sin fallos, lo cual demuestra una buena preparación para escenarios de concurrencia moderada.

Figura 29
Configuración Thread Group (100 hilos 20 segundos)



Nota. Elaboración propia.

Figura 30
Resultados Obtenidos

The screenshot shows the JMeter Summary Report. The report includes columns for Label, # Samples, Average, Min, Max, Std. Dev., Error %, Throughput, Received KB/sec, Sent KB/sec, and Avg. Bytes. The data is as follows:

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
/api/v1/User/login	100	878	476	1973	324.44	0.00%	4.9/sec	4.09	1.32	855.0
Login	100	878	476	1973	324.44	0.00%	4.9/sec	4.09	1.32	855.0
/api/v1/Crop/list	100	190	84	659	108.67	0.00%	5.3/sec	28.42	2.75	5474.0
/api/v1/CultivationSystem/pagination	100	188	94	590	104.38	0.00%	5.4/sec	7.36	2.87	1403.0
/api/v1/Crop/pagination?search=prueba	100	172	89	404	81.65	0.00%	5.4/sec	1.26	2.91	238.0
Dashboard	100	551	272	1185	213.39	0.00%	5.3/sec	36.55	8.35	7115.0
/api/v1/User/update	100	218	100	738	134.79	14.00%	5.5/sec	6.01	7.57	1118.1
Profile	100	218	100	738	134.79	14.00%	5.5/sec	6.01	7.57	1118.1
TOTAL	800	411	84	1973	354.82	3.50%	38.4/sec	85.23	31.04	2272.0

Nota. Elaboración propia.

En la primera prueba, configurada con 20 hilos y un periodo de rampa de 4 segundos, el sistema obtuvo un tiempo de respuesta promedio de 257 ms, con una tasa de error del 0%, lo cual indica una buena capacidad de respuesta ante una carga moderada. En contraste, la segunda prueba se realizó con una carga más exigente, utilizando 100 hilos y un ramp-up de 20 segundos. En este escenario, el tiempo de respuesta promedio aumentó a 411 ms y se presentó un 3.5% de errores, siendo estos principalmente en el servicio de actualización de usuario (/api/v1/User/update), el cual mostró una tasa de error del 14%. Este comportamiento sugiere que, aunque el sistema mantiene un rendimiento aceptable bajo alta concurrencia, existen ciertos servicios que requieren optimización para garantizar estabilidad a gran escala.

2.5. Pruebas de Calidad y Rendimiento (SONARQUBE)

2.5.1. Frontend

El objetivo de estas pruebas fue: Evaluar la calidad del código del modulo de reportes mediante indicadores **SonarQube**, identificando riesgos y definiendo acciones antes de dato como finalización el desarrollo.

El alcance se limite a la carpeta **/src/views/reports**, que corresponde al Módulo de Reportes y el archivo de lógica **/src/services/reportsServices.js**.

Tabla 17
Pruebas SonarQube

	Líneas	Seguridad	Fiabilidad	Mantenibilidad	Puntos Críticos	Cobertura	Duplicaciones
Carpeta							

	Código				Errores de Seguridad		
reports	5,065	0	56	203	0	0.0 %	59.8%
controlledEnvironment	792	0	5	27	0	0.0 %	72.6%
datasets	1,250	0	11	68	0	0.0 %	65.1%
main	773	0	5	24	0	0.0 %	81.3%
modeling	339	0	2	2	0	0.0 %	12.2%
nonControlledEnvironment	732	0	5	25	0	0.0 %	79.1%
semicontrolledEnvironment	1,179	0	28	57	0	0.0 %	32.5%

Nota. Elaboración Propia

El análisis del código fuente mediante la herramienta SonarQube permitió obtener métricas relevantes sobre la calidad del software en desarrollo. En términos de tamaño, el módulo con mayor cantidad de líneas de código (LOC) es reports/ con 5,065 LOC, seguido por datasets/ y semicontrolledEnvironment/. No se detectaron vulnerabilidades críticas de seguridad en ningún módulo, lo que indica que el código no presenta riesgos significativos desde esta perspectiva. Sin embargo, el módulo reports/ presenta una mayor concentración

de problemas de fiabilidad (56) y mantenibilidad (203), evidenciando una alta deuda técnica asociada, principalmente por la lógica compleja y un 59.8 % de código duplicado.

Por otro lado, todas las carpetas muestran un 0 % de cobertura de pruebas, modulo aun sin test, ya que no se implementaron test sobre **SonarQube**, importante resaltar que el módulo main/, con 773 líneas de código, no se considera para el despliegue final, ya que su propósito fue exclusivamente realizar pruebas iniciales del sistema. Por tanto, aunque presenta un nivel considerable de duplicaciones (81.3 %), no se piensa implementar en su totalidad del 100% Este tipo de evaluaciones proporciona un panorama claro de las prioridades en cuanto a calidad del código y esfuerzo de mejora técnica.

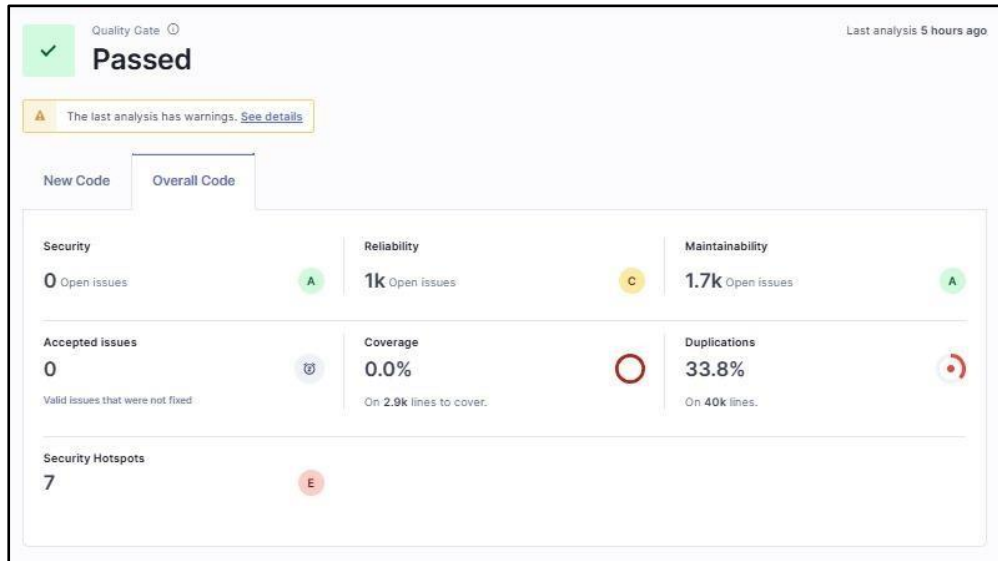
Tabla 18
reportService.js

reportsServices.js	214	0	1	3	0	0.0%	0.0%
--------------------	------------	----------	----------	----------	----------	-------------	-------------

Nota. Elaboración Propia

En el siguiente apartado se muestra la **Rentabilidad en 1** lo que indica un posible escenario de error (validación de entrada nula), por otro el apartado de **Mantenibilidad (3)** indicando que el código cuenta con funciones largas y complejas, **Cobertura (0%)** ya que no se implementaron pruebas sobre la herramienta de **SonarQube** y por último el apartado de duplicados en **(0%)**, indicando implementación limpia del código.

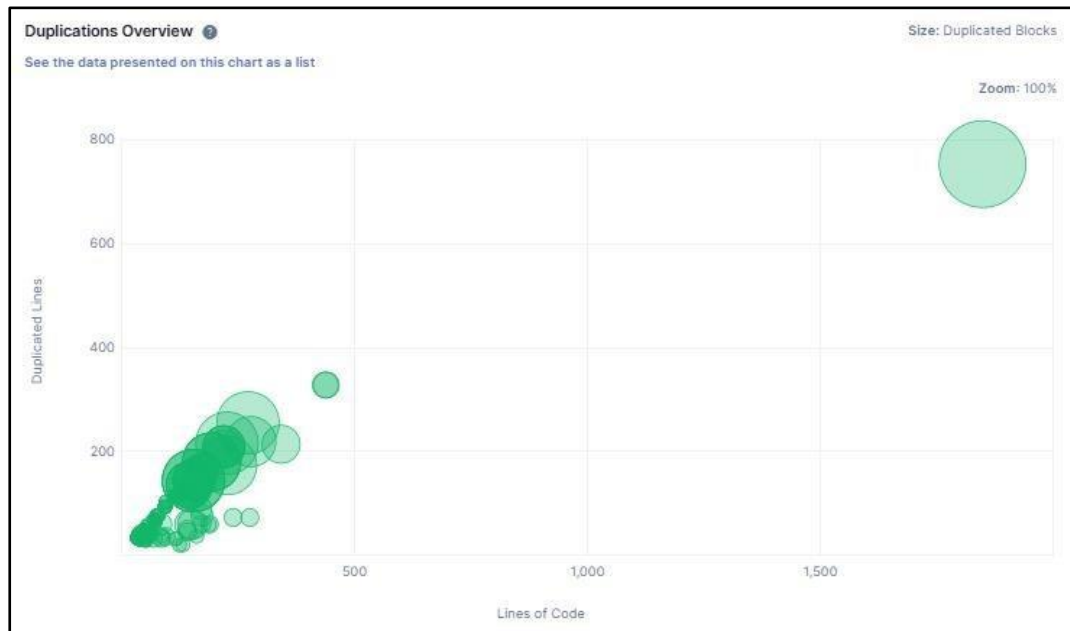
Figura 31
Resultado General de calidad



Nota. Elaboración Propia

En la figura se muestra que, aunque se han identificado algunas advertencias. Los puntos más críticos es el duplicado de código y la cobertura de pruebas que están en 0. No obstante la seguridad no presenta problemas abiertos, y la mantenibilidad, la confiabilidad se consideran en rangos aceptables.

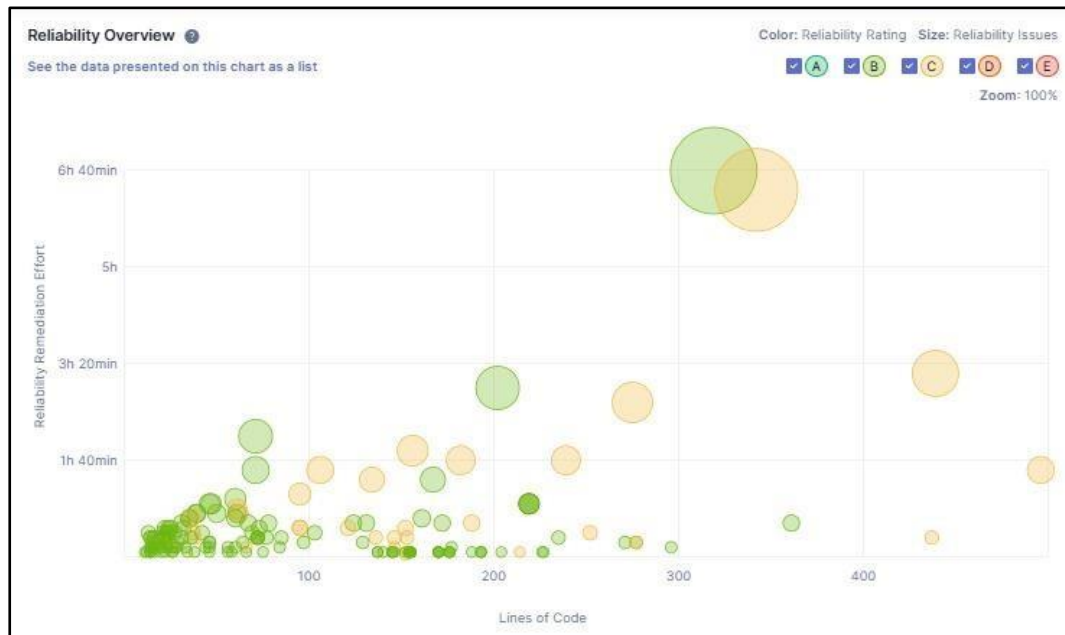
Figura 32
Duplicado de Código



Nota. Elaboración Propia

Esta gráfica muestra los bloques de código duplicado en el proyecto. Se observa que existe una cantidad considerable de duplicación, con un caso particular donde más de 700 líneas están duplicadas en un mismo archivo. Esto puede afectar negativamente la mantenibilidad del código, ya que los cambios deben replicarse en múltiples lugares.

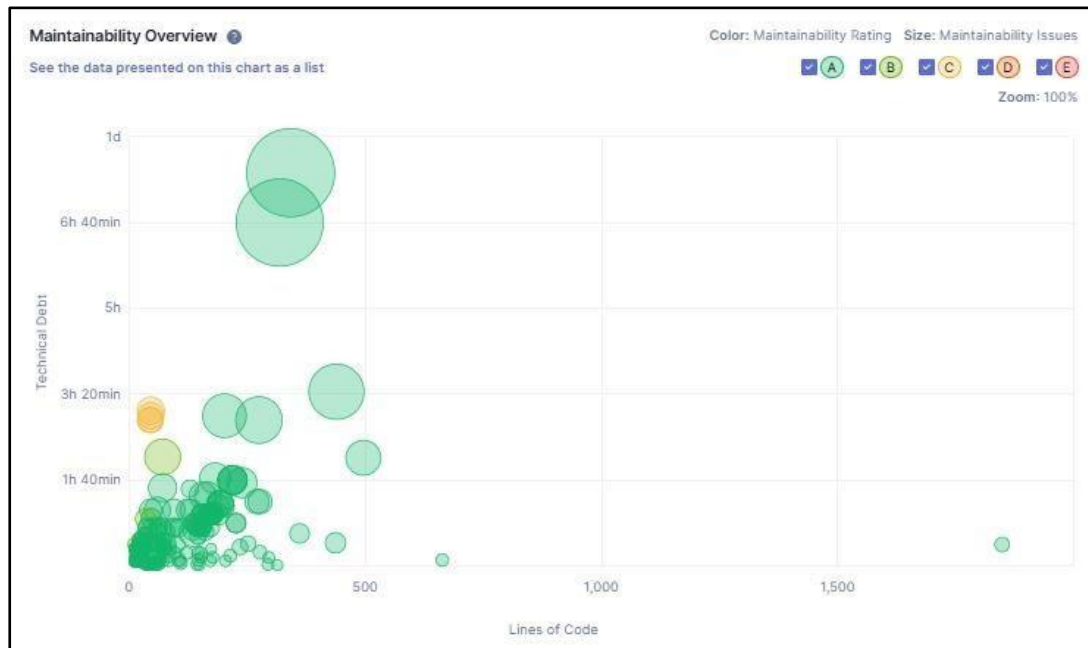
Figura 33
Flexibilidad / Confiabilidad



Nota. Elaboración Propia

Aquí se analiza el esfuerzo necesario para remediar problemas de confiabilidad en el código. La mayoría de los archivos tienen problemas menores, pero algunos módulos más grandes muestran un mayor tiempo de remediación (hasta 6 horas 40 minutos). El color amarillo indica que algunas partes del código tienen una calificación de confiabilidad C.

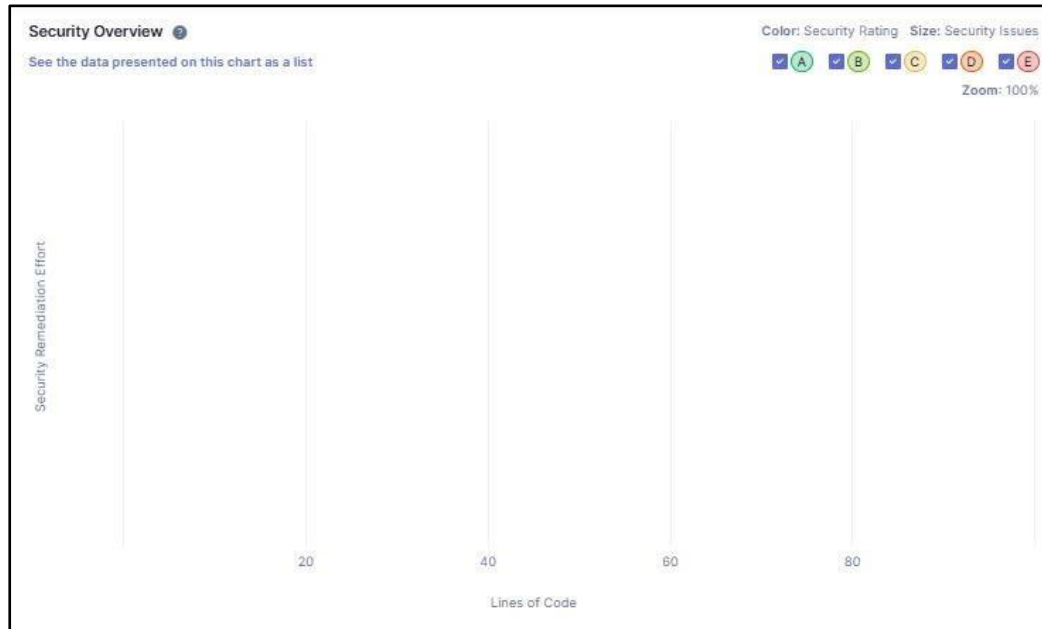
Figura 34
Mantenibilidad



Nota. Elaboración Propia

Esta sección evalúa la deuda técnica relacionada con la mantenibilidad del proyecto. Se observa que la mayoría de los problemas requieren poco tiempo para corregirse, aunque existen unos pocos archivos que podrían necesitar hasta un día completo de trabajo. Afortunadamente, la mayoría de los elementos analizados tienen una buena calificación de mantenibilidad (verde).

Figura 35
Seguridad



Nota. Elaboración Propia

El análisis de seguridad muestra que no hay problemas abiertos en el código en cuanto a vulnerabilidades, pero sí se detectaron 7 "Security Hotspots" (áreas sensibles que podrían convertirse en vulnerabilidades si no se gestionan correctamente). No se requiere un esfuerzo inmediato de remediación en términos de líneas de código afectadas.

En conclusión, el análisis realizado no evidenció riesgos de seguridad, ya que SonarQube no reportó vulnerabilidades ni hotspots en el código evaluado. Es importante destacar que este informe se enfoca exclusivamente en el módulo de **Reportes**, por lo que los resultados presentados corresponden únicamente a esta sección del sistema.

2.5.2. Backend

No Aplica ya que el proyecto no cuenta con requerimientos de back End

2.6. Estimación por Puntos de Función

Lista y clasificación de unidades funcionales

En esta sección, se definen las unidades funcionales para cada uno de los requerimientos planteados en el documento IEEE. Estas unidades permitirán aplicar la estimación por puntos de función, además de asignarles un tipo y nivel de complejidad.

Identificación de los elementos funcionales

Se dividen en cinco tipos:

- **Entradas Externas (EE):** Datos ingresados por el usuario al sistema.
- **Salidas Externas (SE):** Información procesada y mostrada al usuario.
- **Consultas Externas (CE):** Combinación de entrada y salida sin procesamiento complejo.
- **Archivos Lógicos Internos (ALI):** Bases de datos o archivos que pertenecen al sistema.
- **Archivos de Interfaz Externa (AIE):** Archivos compartidos con otros sistemas.

Asignación de peso a cada elemento

Se asigna una ponderación dependiendo la complejidad del elemento, basado en la **tabla del international function pont users Group (IFPUG).**

Tabla 19

Pesos de asignación según international function pont users Group (IFPUG)

Elemento	Simple	Medio	Complejo
EE	3	4	6

SE	4	5	7
CE	3	4	6
ALI	7	10	15
AIE	5	7	10

Nota. International Organization for Standardization. (2009). ISO/IEC 20926:2009 - Software and systems engineering — Software measurement — IFPUG functional size measurement method. International Organization for Standardization.

En la tabla 19 se describe el peso asignado a cada una de las categorías a la que se idéntica cada elemento funcional basándose en lo que proporciona el documento de la norma ISO.

Tabla 20
Identificación y clasificación de unidades funcionales

ID	Requerimiento	Tipo de funcionalidad	Complejidad	PF asignados
RF-01	Interfaz Web	Entrada Externa (EI)	Bajo	3
RF-02	Visualización de Datasets	Consulta Externa (EQ)	Medio	4
RF-03	Visualización de metodología y técnicas	Consulta Externa (EQ)	Medio	4
RF-04	Visualizar modelo matemático	Consulta Externa (EQ)	Alto	6

6RF-05	Visualizar técnica de análisis de imágenes	Consulta Externa (EQ)	Medio	4
R4F-06	Generación de reportes	Salida Externa (EO)	Alto	7
Total UFPC:				28

Nota. Elaboración propia.

En la tabla 20 se muestra cada uno de los requerimientos funcionales del proyecto, la categoría a la que pertenece según el elemento funcional teniendo presente su nivel de complejidad y puntaje asignado.

Tabla 21
Cálculo de grado de influencia

Características Generales del Sistema	Grado de influencia
Comunicación de datos	3
Procesamiento de datos distribuidos	3
Rendimiento	3
Configuración de uso intensivo	3
Tasa de transacción	4
Entrada de datos en línea	4
Eficiencia del usuario final	4

Actualización en línea	3
Procesamiento complejo	3
Reusabilidad	3
Facilidad de instalación	4
Facilidad operativa	4
Sitios múltiples	3
Facilitar el cambio	3
Total Grado de Influencia (TDI)	47
Valor de Factor de Ajuste (VAF)	1.12

*Nota. Para el cálculo del valor del factor de ajuste se tuvo en cuenta la fórmula presentada en el documento, siendo la siguiente: $VAF = (TDI * 0.01) + 0.65$. Autoría propia.*

En la tabla 21 se muestra el cálculo de grado de influencia para cada uno de los aspectos mencionados, además se muestra el cálculo del valor de factor de ajuste el cual es crucial para el cálculo de los siguientes componentes.

- **Cálculo de puntos de función ajustados**

Para calcular los puntos de función ajustados, se empleó la siguiente fórmula tomada del documento ISO 20926. En esta fórmula, UFPC representa la identificación y clasificación de unidades funcionales, mientras que VAF corresponde al cálculo del grado de influencia.

$$FPC = UFPC * VAF$$

$$FPC = 28 * 1.12$$

$$FPC = 31.36$$

- **Cálculo de estimación**

La estimación del esfuerzo considera el número de horas que una persona requiere para completar un punto de función, representado como HFP. Para calcularlo, se tomó como referencia el esfuerzo realizado por un individuo en función del tiempo asignado al proyecto, que es de **10 meses**, equivalentes a **7,200 horas**.

Una vez determinado este tiempo, se procede a calcular la tasa de entrega utilizando la siguiente fórmula:

$$TDE = \frac{HORAS}{PUNTOS DE FUNCION}$$

$$TDE = \frac{7200}{31.36}$$

$$TDE = 229,592$$

- **Estimación de Costo**

La estimación de costos considera la nómina mensual definida por el equipo de desarrolladores del proyecto, representada como **MP**, y el número de integrantes del equipo, representado como **NE**. Además, se toma en cuenta la estimación de tiempo mensual obtenida previamente. Con estos valores, se aplicó la siguiente fórmula para realizar el cálculo:

$$CE = MP * NE * TDE_{Mensual}$$

$$CE = \$ 2.500.000 * 2 * 2.295$$

$$CE = \$ 11.475.000_{Mensualmente}$$

$$\text{Total Proyecto} = \text{Cant. Meses} * \text{cant pago recurso}$$

$$\text{Total Proyecto} = \$ 114.750.000$$

2.7. Integración de Datos

En un principio, se contempló utilizar únicamente datos reales para modelar el crecimiento de las plantas. La idea era contar con un registro completo que incluyera todas las etapas del desarrollo, junto con las variables ambientales registradas en cada momento. Sin embargo, debido a factores externos al proyecto, no se logró reunir toda la información necesaria, por lo que fue preciso recurrir a la generación de datos sintéticos de cada una de las variables ambientales consideradas para completar las series temporales faltantes.

2.7.1. Ambiente y Plantas

El prototipo computacional para el manejo de las plantas cuenta con 3 ambientes con diferentes características. El primero de estos ambientes es uno **semicontrolado** donde hay 3 plantas y 2 grupos de sensores, grupos conformado por un sensor capacitivo de humedad, una fotorresistencia, un DHT22 y un sensor de UV. Este primer ambiente se considera **semicontrolado** porque a pesar de tener una fuente de luz artificial, también recibe luz solar y flujo de aire del exterior. El segundo ambiente es uno controlado, que cuenta con una planta y un grupo de sensores. Este ambiente recibe luz solar en muy poca medida y se encuentra en su mayoría cerrado por lo que no hay mucho flujo de aire

exterior. Y finalmente el tercer ambiente, este se encuentra abierto y con incidencia casi total de luz solar.

Para los capítulos a continuación se tuvo en cuenta el ambiente semicontrolado, con información del ambiente mediante los dos grupos de sensores y de la altura de las 3 plantas correspondientes a este ambiente. Por razones exteriores al proyecto no fue posible contar con información de los demás ambientes, pero en un futuro esos datos estarán disponibles para nuevas investigaciones.

2.7.2. Generación de Datos Sintéticos

Para crear los datos sintéticos necesarios para el modelo, se optó como principal opción el modelo ARIMA o SARIMA dado a su capacidad de captar patrones y tendencias estadísticas con pocos datos como se menciona más adelante en el apartado donde se profundiza en estos modelos. Por otro lado, si la cantidad de datos no llega a ser suficiente para estos modelos anteriormente mencionados, se optará por buscar otras alternativas tales como redes neuronales dado a su capacidad de poder adaptarse a patrones propios de los datos es decir hallar la forma en que se comportan los datos, para lograr esto se pretende hacerlo definiendo una estructura, función de activación, épocas, datos de entrenamiento y salida que se requiera para este proceso.

Generación de Datos Sintéticos con una Pequeña Muestra

La generación de datos sintéticos con una pequeña muestra es el proceso de crear datos artificiales que imitan las propiedades estadísticas y estructurales de un conjunto de datos real, pero partiendo de un número limitado de observaciones. Este enfoque es útil cuando los datos reales son escasos o sensibles, ya que permite entrenar modelos de

inteligencia artificial sin comprometer la privacidad ni la integridad de la información original (Arroni del Riego, 2024).

Para lograrlo, se emplean técnicas como las Redes Generativas Antagónicas (GANs), los Autoencoders Variacionales (VAEs), los modelos de difusión y las redes neuronales recurrentes (RNN). Estas metodologías permiten aprender la distribución subyacente de los datos y generar nuevas instancias que siguen patrones similares a los datos reales (Esteban Lasso et al., 2023).

Estudios Sobre Generación de Datos Sintéticos con Muestras Pequeñas

1) Generación de datos sintéticos tabulares con modelos de difusión

Un estudio reciente propuso un modelo basado en redes de difusión para generar datos sintéticos tabulares utilizando muestras de entre 10 y 20,000 registros. Se demostró que, incluso con tan solo 10 datos de entrenamiento, el modelo fue capaz de generar datos sintéticos útiles para entrenamiento de modelos de aprendizaje automático (Arroni del Riego, 2024).

2) Generación de imágenes sintéticas para entrenar redes neuronales

En el ámbito de la visión por computadora, se han utilizado redes neuronales convolucionales (CNN), como Mask R-CNN, para generar imágenes sintéticas de objetos de cocina a partir de fondos y segmentaciones de objetos reales. Esta técnica permitió entrenar modelos con conjuntos de datos generados artificialmente sin necesidad de grandes volúmenes de datos reales (Benages-Pardo et al., 2019).

3) Generación de datos médicos sintéticos con GANs y VAEs

Se han desarrollado modelos como TVAE (Tabular Variational Autoencoder) y CTGAN (Conditional Tabular GAN) para generar datos sintéticos en entornos médicos. Estos modelos logran preservar la estructura y relaciones de los datos originales, asegurando su utilidad en la investigación médica sin comprometer información confidencial (Esteban Lasso et al., 2023).

4) Generación de datos sintéticos de conducción con TimeGAN

En el sector automotriz, se ha utilizado TimeGAN, un modelo híbrido de redes neuronales recurrentes (RNN) y GANs, para generar datos sintéticos de eventos de conducción. Este modelo captura la distribución temporal y espacial de los datos, permitiendo el entrenamiento de sistemas de conducción autónoma sin necesidad de grandes volúmenes de datos reales (Tamayo Urgilés, 2023).

2.7.3. Modelo ARIMA y SARIMA

El modelo ARIMA (Autoregressive Integrated Moving Average) es una técnica estadística utilizada para el análisis y la predicción de series de tiempo. Este modelo combina tres componentes principales: autorregresión (AR), que incorpora la relación entre observaciones pasadas; integración (I), que diferencia la serie para hacerla estacionaria; y media móvil (MA), que modela la dependencia entre valores residuales de la serie (Box et al., 2015).

Matemáticamente, un modelo ARIMA se representa como $ARIMA(p, d, q)$, donde p es el orden del componente autorregresivo, d es el número de diferencias necesarias para hacer la serie estacionaria, y q es el orden del componente de media móvil (Hyndman & Athanasopoulos, 2018). Este modelo es ampliamente utilizado en la predicción de datos

económicos, financieros y científicos debido a su capacidad para capturar patrones temporales en los datos (Wei, 2006).

En este proyecto, la elección del modelo ARIMA estuvo motivada principalmente por la limitada cantidad de datos reales disponibles sobre variables ambientales como la humedad relativa, temperatura y luminosidad. Debido a que estos datos se presentaron en forma de series temporales incompletas, era esencial utilizar un método que pudiera identificar patrones claros en pocos puntos de referencia. ARIMA se adaptó particularmente bien a estas condiciones, permitiendo generar datos sintéticos cercanos a la realidad conservando la estructura original observada en los datos parciales. Por otro lado, se conoce que las variables ambientales siguen un patrón cíclico diario, resultando más útil usar la extensión del modelo SARIMA que está diseñado para capturar ese componente estacional (Pongdatu y Putra, 2018).

2.7.4. Estructura Temporal

Antes de generar los datos fue necesario definir una serie temporal que simule los instantes en que se hubieran hecho las mediciones reales. En este ejercicio, tomando como base la definición de recolectar el conjunto de datos de los sensores reales cada 10 minutos, también se siguió esta misma frecuencia, aunque no sea la misma frecuencia usada en el desarrollo del modelo matemático.

Por otro lado, una de las finalidades del estudio realizado fue encontrar relaciones entre las variables ambientales y la altura, no obstante, el tiempo resulta siendo una variable con una gran importancia en el crecimiento, por lo cual, fue crucial expresar el tiempo de tal manera permitiera integrar la variable en los modelos y no generara sesgos al momento de usar fechas específicas. Es decir, representar el paso del tiempo, no las fechas en un

formato natural. Para lograr esto se tomó la decisión de trabajar con fechas relativas, inicialmente usando la misma unidad que el tiempo de muestreo, es decir, minutos, y tomar como punto de partida para las mediciones sintéticas 3 meses después de las plantas haber sido sembradas, pero siendo la primera marca el momento en el que se realizó la siembra original.

Lo dicho hasta aquí supone que la siembra original fue el 1 de noviembre de 2023; lo cual sería nuestro minuto 0, el inicio del estudio el 1 de febrero de 2024; tratándose del minuto 220320 (ver Tabla 20), es decir 92 días después y la finalización el 26 de marzo.

Tabla 22

Fecha y minutos transcurridos desde la siembra.

Fecha	Minutos desde la siembra
2023-11-01 05:00	0
2023-11-01 05:10	10
2024-02-01 05:00	132480
2024-02-01 05:10	132490

Nota. Frecuencia de muestreo = 10 minutos. El 1 de febrero de 2024 inicia cómo tal el estudio, siendo el minuto 132480 después de la siembra.

Esta transformación a tiempos relativos permite integrar la variable tiempo en los modelos sin introducir sesgos derivados del uso de fechas absolutas, facilitando así la identificación de patrones de crecimiento y relaciones con las variables ambientales.

```

from datetime import datetime, timedelta
import csv
inicio = datetime(2023, 11, 1, 5, 0)
final = datetime(2025, 3, 26, 5, 0)
# Función para calcular los minutos transcurridos desde la fecha inicial de la siembra
def minutos_transcurridos(fecha_ingresada):
    diferencia = fecha_ingresada - inicio
    minutos = int(diferencia.total_seconds() / 60) # Convertir la diferencia a minutos
    return (minutos // 10) * 10 # Redondear al múltiplo de 10 más cercano
# Función para generar los registros de fecha, hora y minutos transcurridos
def generar_registros(inicio, final):
    registros = []
    fecha_actual = inicio
    while fecha_actual <= final:
        minuto = minutos_transcurridos(fecha_actual)
        registros.append([fecha_actual.strftime('%Y-%m-%d %H:%M'), minuto])
        fecha_actual += timedelta(minutes=10)
    return registros

def guardar_csv(registros, nombre_archivo="registros_temporales.csv"):
    with open(nombre_archivo, mode="w", newline="") as archivo:

```

```

    escritor = csv.writer(archivo)
    escritor.writerow(["Fecha_Hora", "Minutos_Desde_Siembra"]) # Encabezados
    escritor.writerows(registros) # Escribir todos los registros
# Ejecución principal del script
if __name__ == "__main__":
    registros = generar_registros(inicio, final)
    guardar_csv(registros)

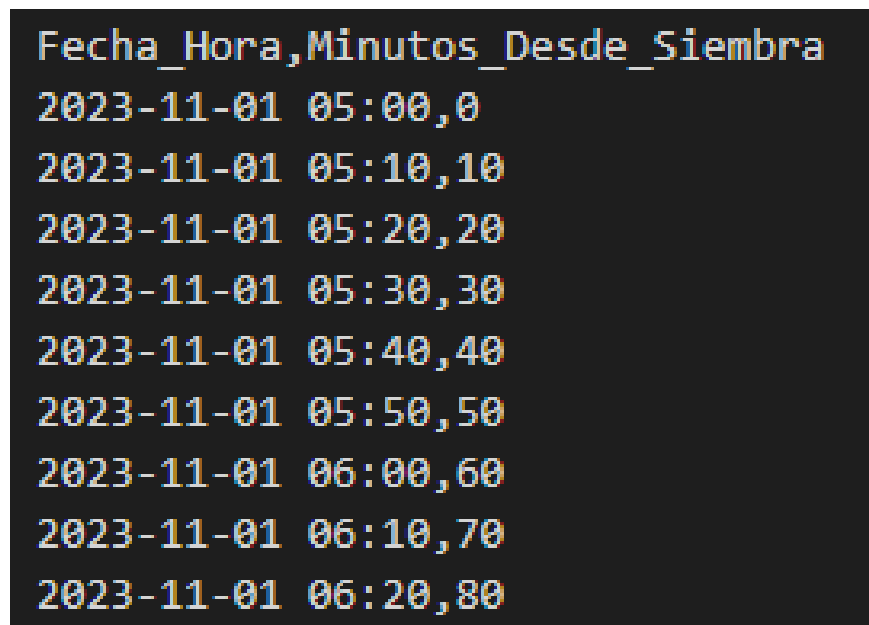
    print(f"Se ha generado el archivo 'registros_temporales.csv' con {len(registros)}
registros.")

```

Generando un archivo CSV con la siguiente estructura:

Figura 36

Estructura archivo CSV con fechas y minutos desde la siembra inicial.



```

Fecha_Hora,Minutos_Desde_Siembra
2023-11-01 05:00,0
2023-11-01 05:10,10
2023-11-01 05:20,20
2023-11-01 05:30,30
2023-11-01 05:40,40
2023-11-01 05:50,50
2023-11-01 06:00,60
2023-11-01 06:10,70
2023-11-01 06:20,80

```

Nota. En la imagen se pueden ver las dos columnas del archivo CSV de fechas y minutos desde la siembra y algunos registros. Elaboración propia.

2.7.5. Preprocesamiento Datos Reales

Los datos reales se obtuvieron en formato CSV conteniendo estas columnas: timestamp, tempdht, humdht, humedsuelo_porcentaje, uv y lum. Estos datos corresponden

al lapso entre el **17 de marzo de 2025 a las 8:42 p. m hasta el día 20 de marzo de 2025 a las 7:37 p. m.** Al realizar una revisión sobre la totalidad de los registros, se encontraron las siguientes observaciones:

- Hay mediciones con valores nulos.
- Se sabe que el servidor que almacenaba los datos contaba con horario UTC, por lo que siempre iba a estar 5 horas delante del instante de la medición real.

Para reparar los valores nulos se hizo uso de la interpolación lineal y adicionalmente se retrasó la variable timestamp 5 horas que ajustarla al horario local usando el siguiente script con ayuda de la librería Pandas:

```
import pandas as pd
# Cargar archivo CSV
df = pd.read_csv('resultado_group2.csv')
# Convertir 'timestamp' a datetime
df['timestamp'] = pd.to_datetime(df['timestamp'])
# Restar 5 horas por diferencia de horario
df['timestamp'] = df['timestamp'] - pd.Timedelta(hours=5)
# Interpolación lineal para rellenar valores faltantes
df.interpolate(method='linear', inplace=True)
# Guardar nuevamente CSV
df.to_csv('resultado_group2_completo.csv', index=False)
print("Archivo guardado: resultado_group2_completo.csv")
```

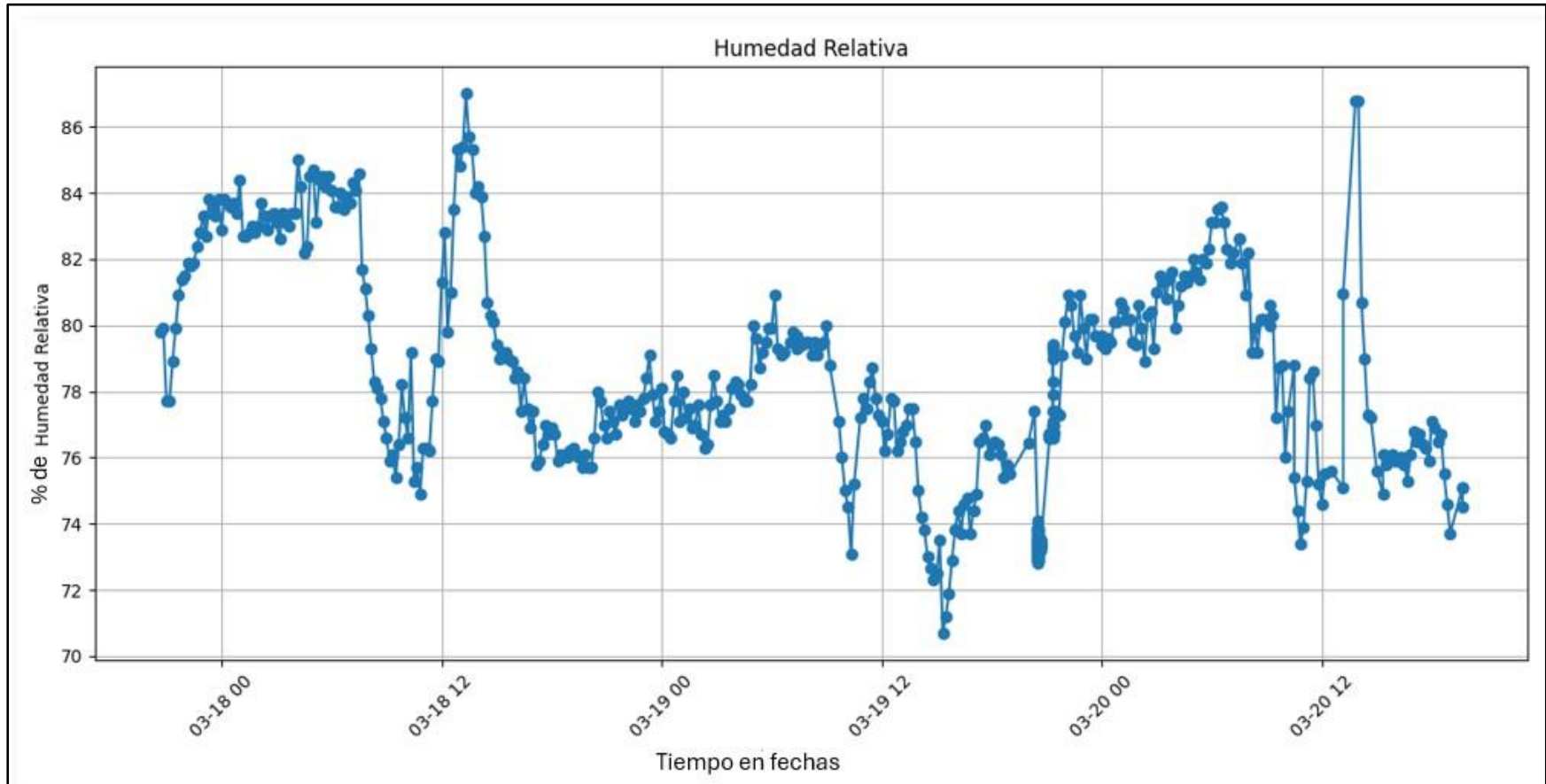
La interpolación lineal es un método matemático utilizado para estimar e insertar un valor desconocido que se encuentra entre dos observaciones conocidas dentro de una serie temporal. Consiste en suponer que el cambio entre estos puntos sigue una tendencia lineal, trazando una recta entre el valor anterior y el posterior para inferir el dato faltante. En nuestro contexto, esta técnica se usó para rellenar valores faltantes (nulos), ya que permite preservar la continuidad del registro temporal basándose únicamente en la información

existente. Tratándose alrededor de un 1% de datos nulos fue conveniente usar esta técnica dada su simplicidad.

Adicionalmente también fue necesario ajustar los minutos de las mediciones reales para que se ajusten partiendo desde el minuto 0, es decir minuto 0, minuto 10, minuto 20, etc. Al finalizar esta primera etapa de preprocesamiento, así se vería la distribución en el tiempo de cada una de las variables.

Figura 37

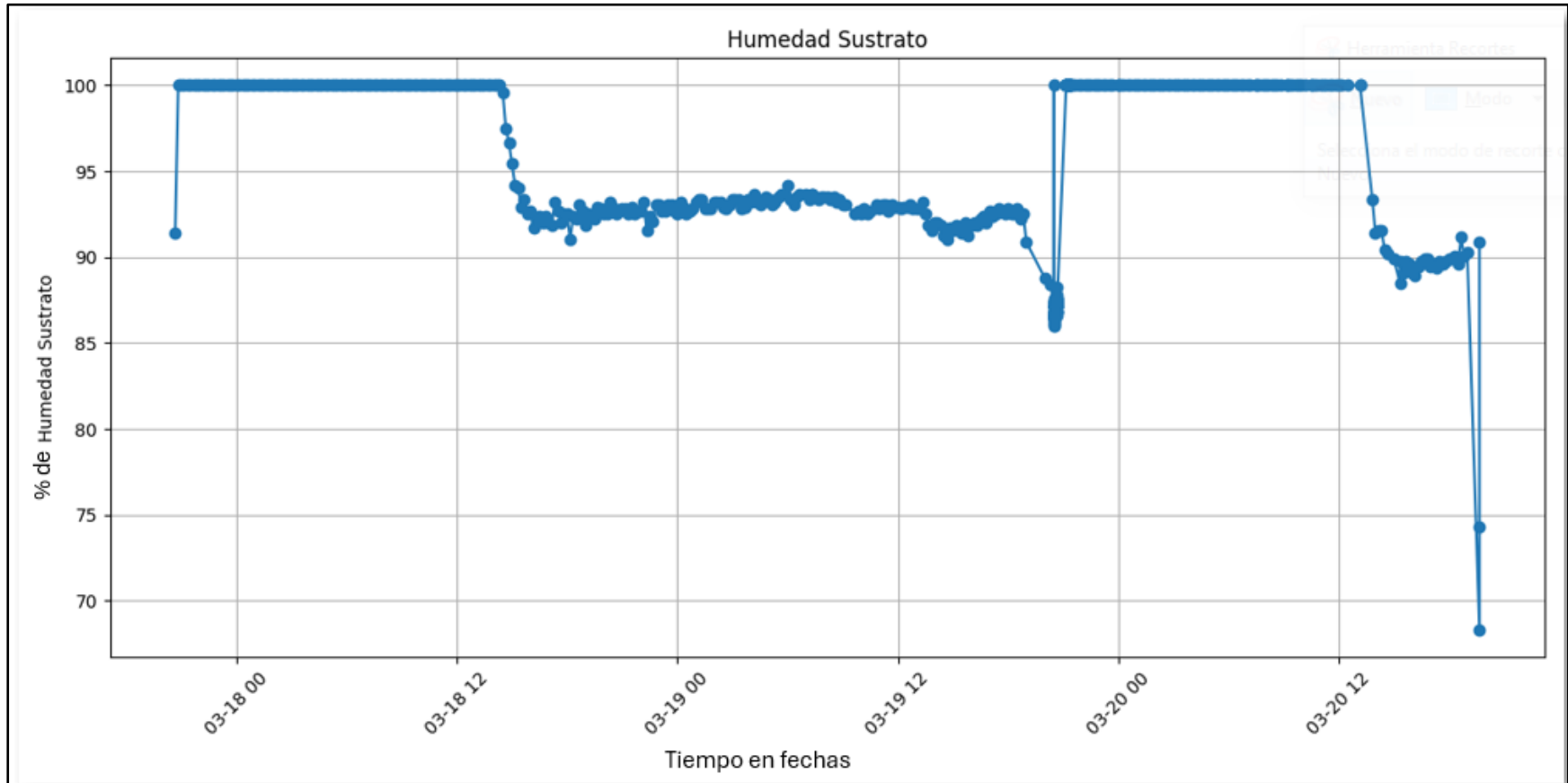
Grafica de porcentaje de humedad relativa vs tiempo del grupo de sensores 1 ambiente semi controlado.



Nota. Se pueden ver los datos del porcentaje de la humedad relativa para el grupo de sensores 1 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Figura 38

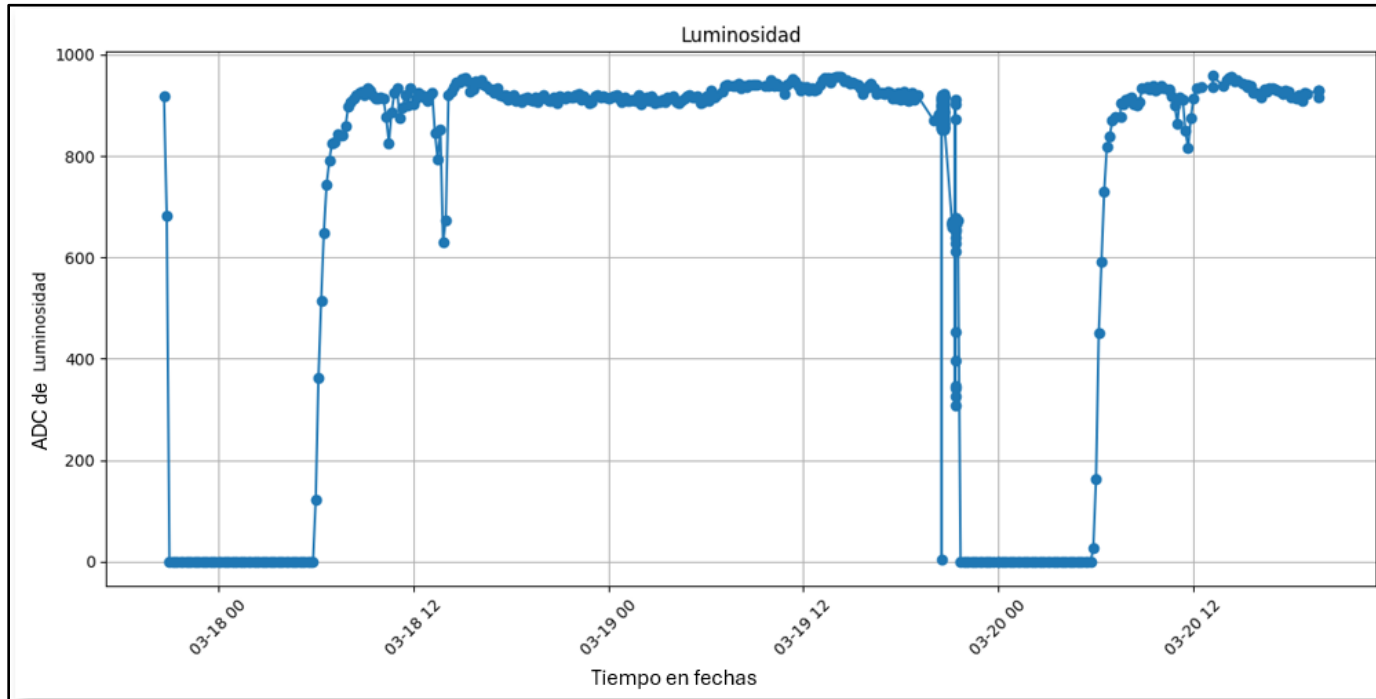
Porcentaje humedad del sustrato vs tiempo del grupo de sensores 1 ambiente semi controlado.



Nota. Se pueden ver los datos del porcentaje de humedad del sustrato vs el paso del tiempo para el grupo de sensores 1 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Figura 39

Datos ADC de Luminosidad vs tiempo real para el grupo de sensores 1 ambiente semi controlado.

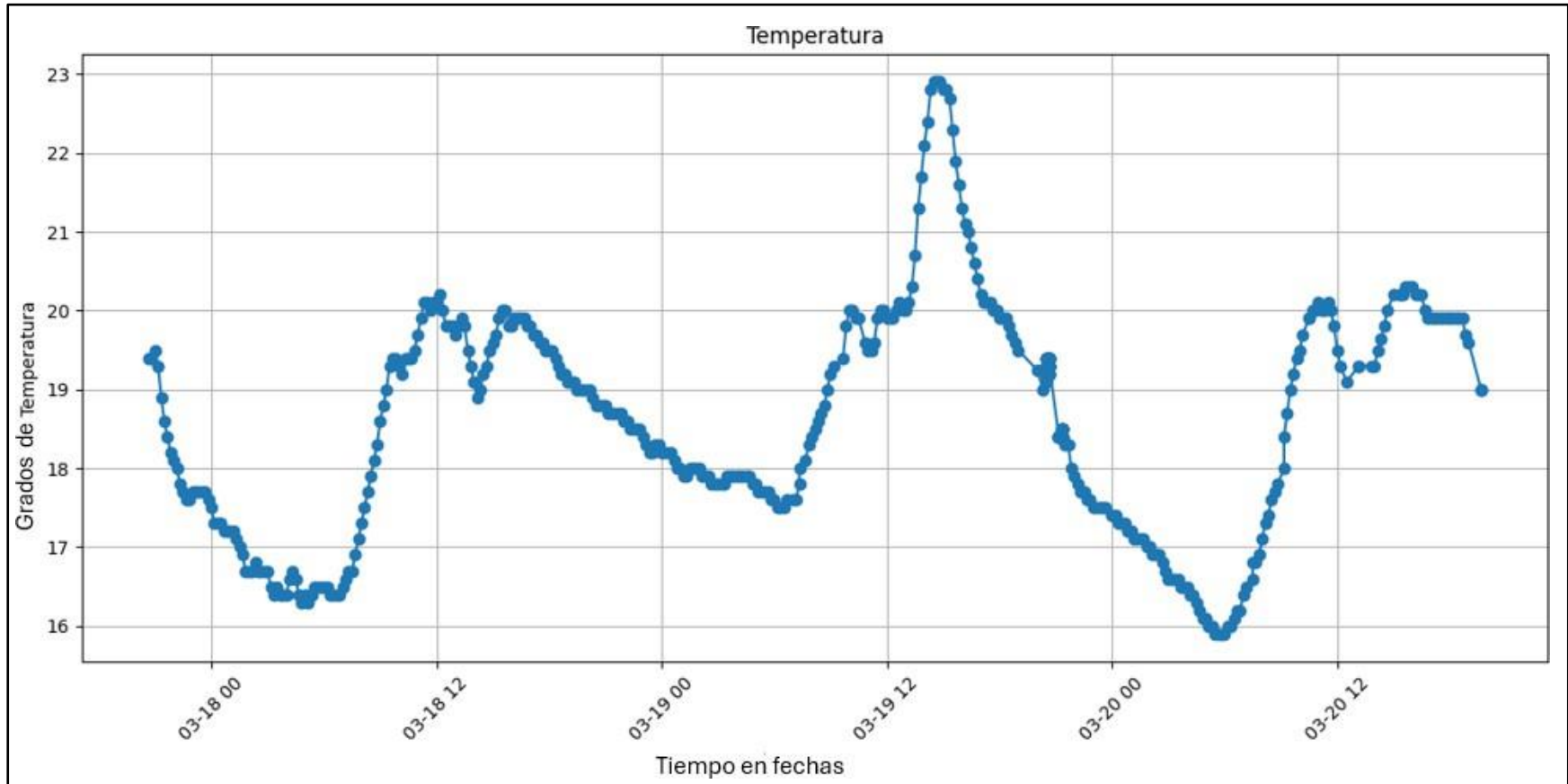


Nota. Se pueden ver los datos ADC de luminosidad vs el tiempo para el grupo de sensores 1 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Inicialmente se consideró que la luminosidad obtenida era en lx, pero realmente se trataba de la salida ADC de los sensores, de la misma manera al esta variable no tener una relevancia en el modelo, no se reconsidero si la medida tuviera importancia o no

Figura 40

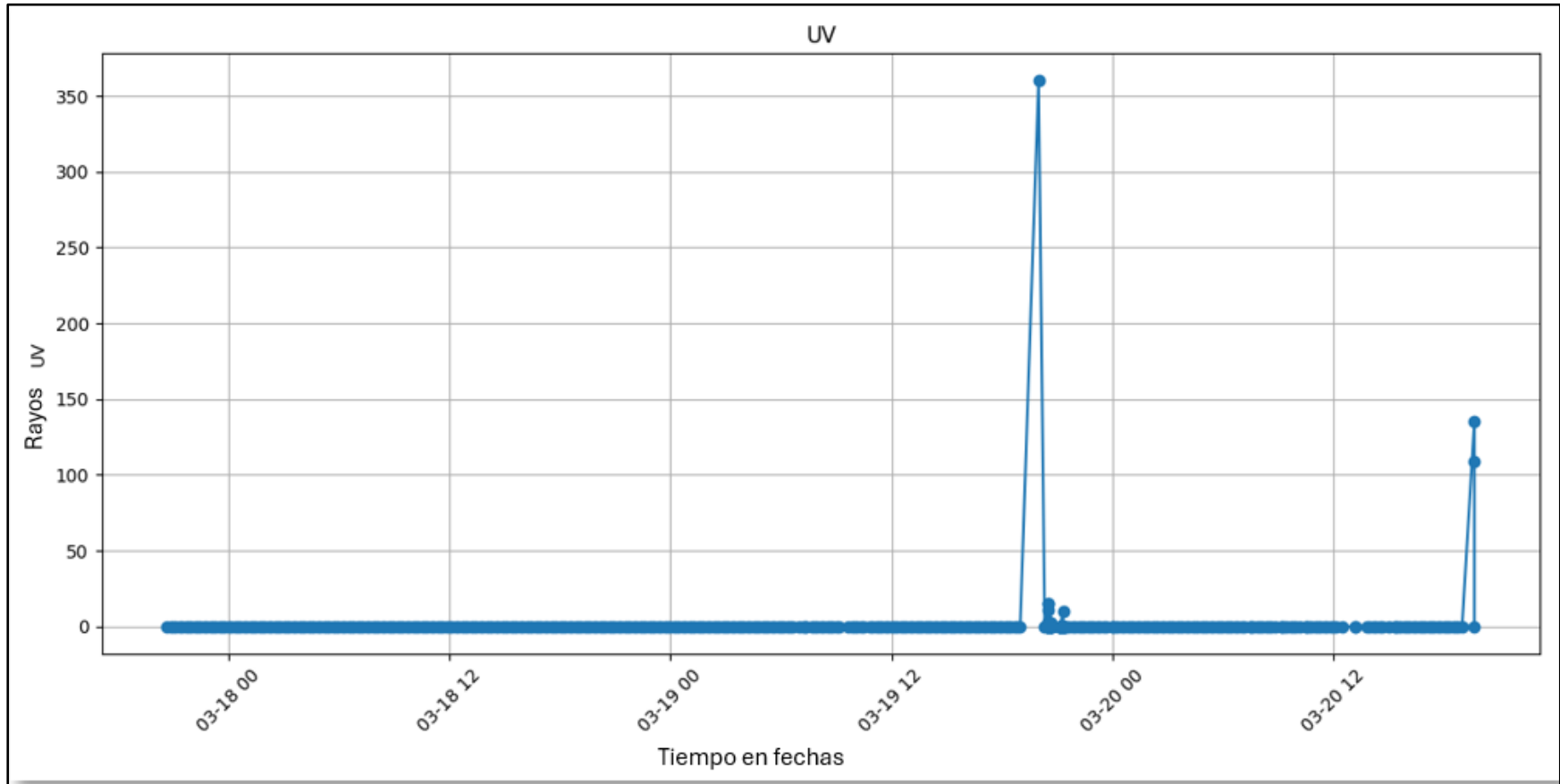
Datos de grados temperatura real vs tiempo para el grupo de sensores 1 ambiente semi controlado.



Nota. Se pueden ver los datos de los grados de temperatura respecto al tiempo para el grupo de sensores 1 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Figura 41

Datos de rayos UV vs el tiempo para el grupo de sensores 1 ambiente semi controlado.

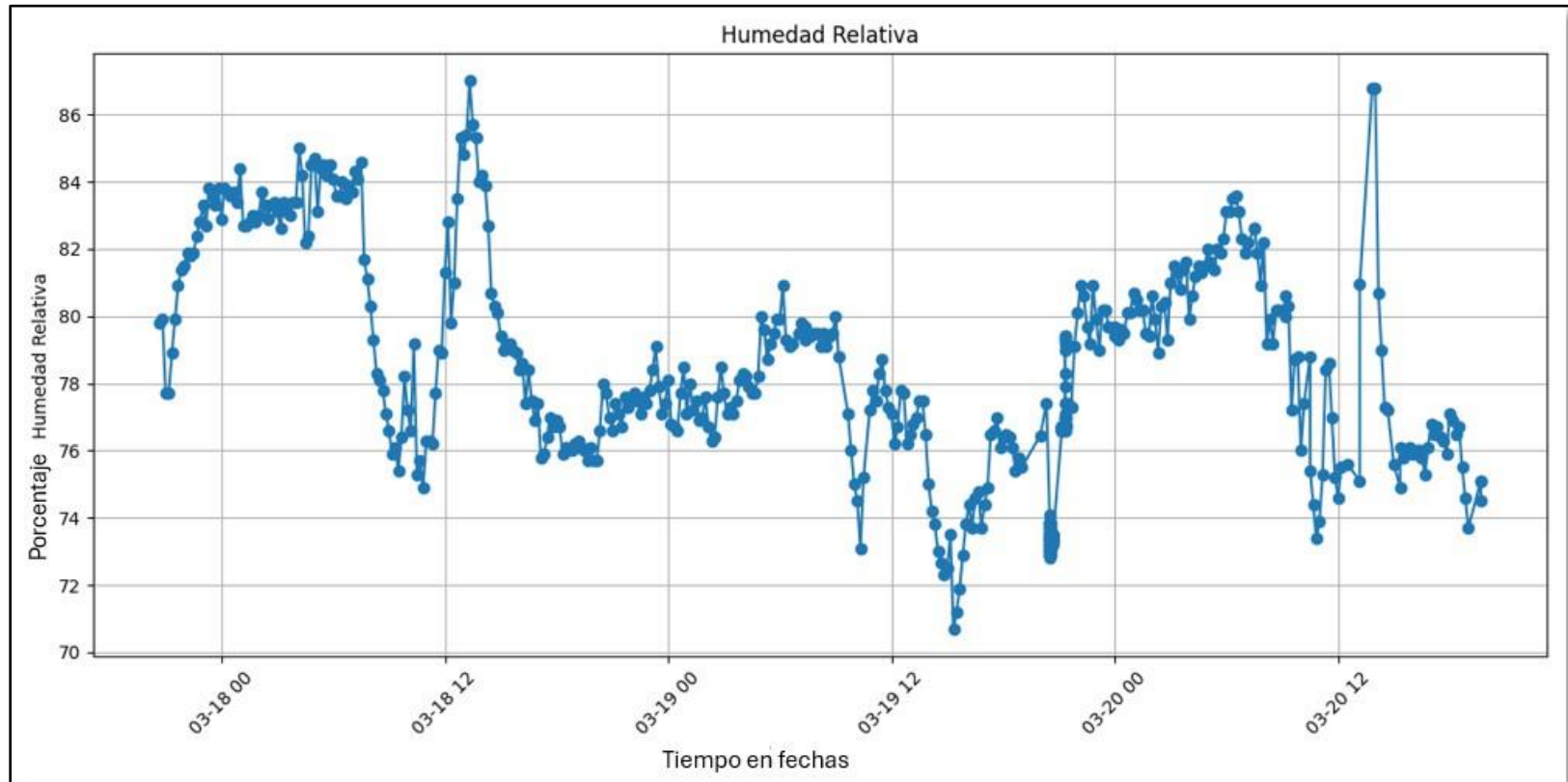


Nota. Se pueden ver los datos de rayos ultravioleta obtenidos con paso del tiempo para el grupo de sensores 1 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Respecto al grupo de sensores 2 del ambiente semi controlado se tiene la información que se encuentra a continuación.

Figura 42

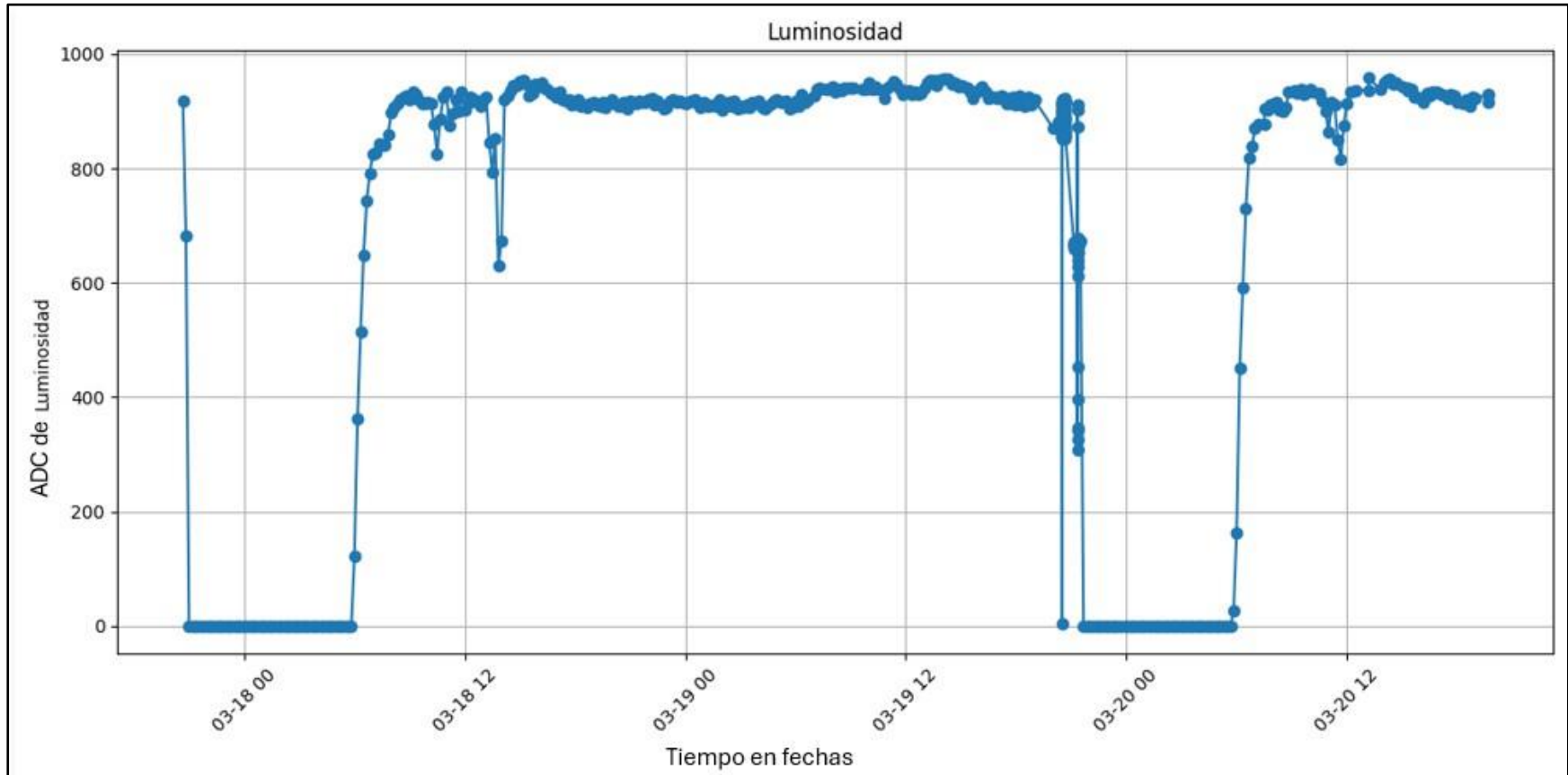
Datos del porcentaje de Humedad relativa vs el tiempo para el grupo de sensores 2 ambiente semi controlado.



Nota. Se pueden ver los datos del porcentaje de humedad relativa con forme el paso del tiempo para el grupo de sensores 2 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Figura 44

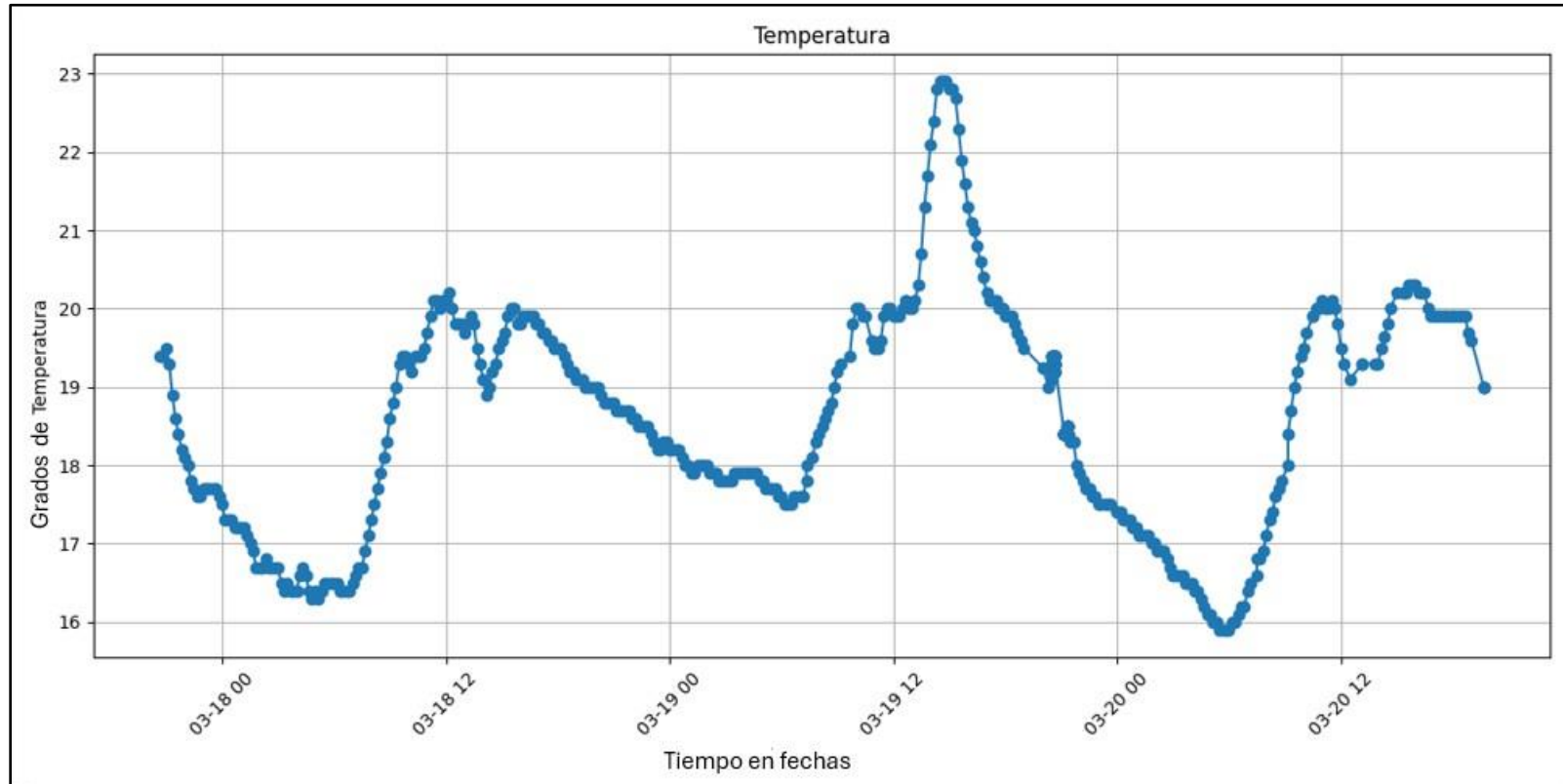
Datos ADC de Luminosidad vs el tiempo para el grupo de sensores 2 ambiente semi controlado.



Nota. Se pueden ver los datos ADC de luminosidad para el grupo de sensores 2 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Figura 45

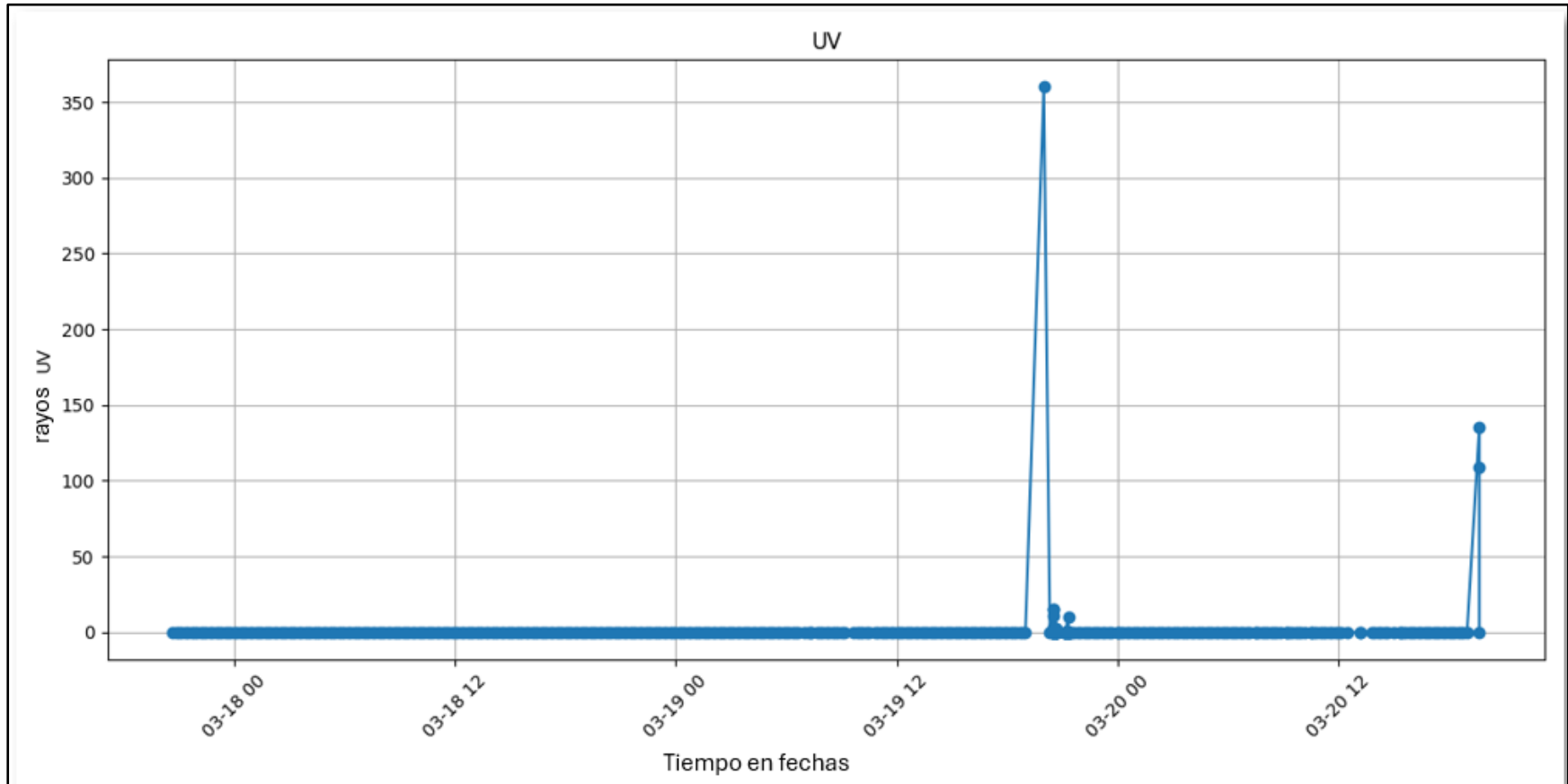
Datos de grados de temperatura vs el tiempo para el grupo de sensores 2 ambiente semi controlado.



Nota. Se pueden ver los datos sobre los grados de temperatura respecto el tiempo para el grupo de sensores 2 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Figura 46

Datos de rayos UV vs el tiempo real para el grupo de sensores 2 ambiente semi controlado.



Nota. Se pueden ver los datos de rayos ultravioleta vs el tiempo para el grupo de sensores 2 del ambiente semi controlado entre el 17 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

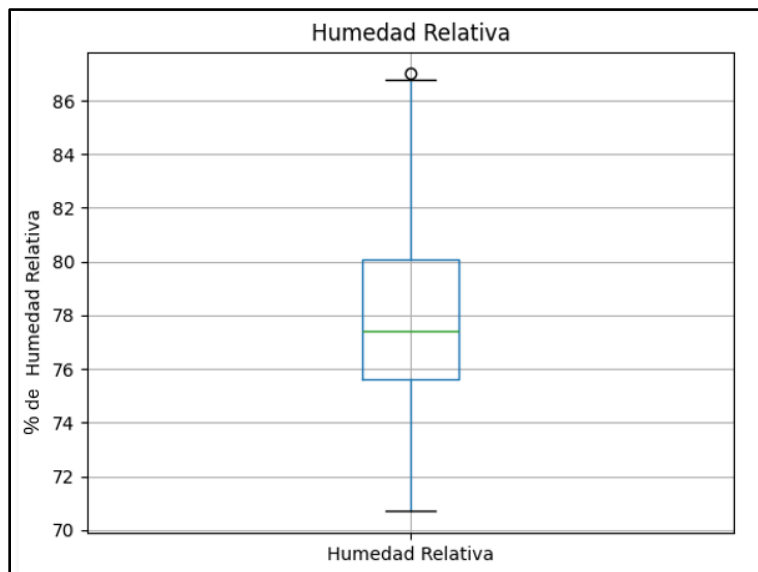
Adicionalmente, en las anteriores figuras se puede observar que los datos siguen un ciclo de día y noche, por lo que es apropiado usar un modelo que pueda captar ciclos estacionarios como SARIMA.

Para la segunda etapa del preprocesamiento de estos datos se analizó si existían suficientes datos outlier como para implementar alguna técnica que los corrigiera. Para esto se graficaron los siguientes histogramas y se consideró la naturaleza de los datos dado el ambiente en que se encuentran.

Como se puede ver en la figura 42 (humedad relativa) la humedad relativa no cuenta con un patrón particularmente extraño, se mantiene en un rango entre 70% y aproximadamente 85% sin muchos valores atípicos como se puede ver en las siguientes figuras.

Figura 47

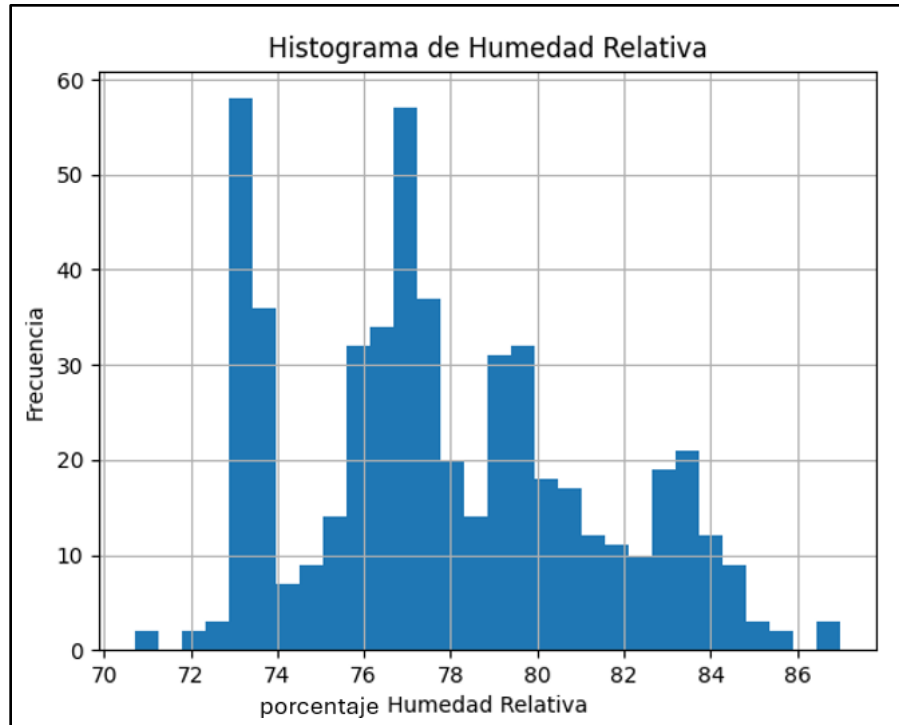
Diagrama de caja sobre los datos de porcentaje de humedad relativa real para el grupo de sensores 1 ambiente semi controlado.



Nota. Diagrama de caja para porcentaje de la variable humedad relativa del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

Figura 48

Histograma porcentaje de humedad relativa grupo de sensores 1 ambiente semi controlado.



Nota. Histograma para porcentaje de la variable humedad relativa del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

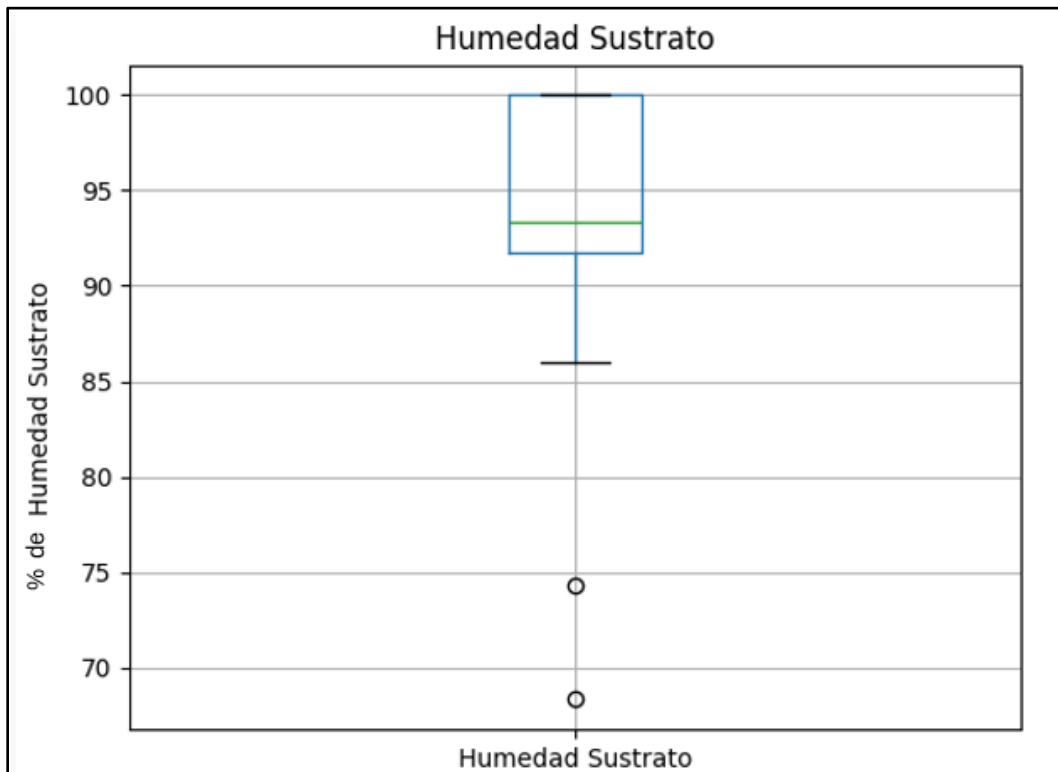
En la figura anterior se puede ver que no existen muchos grupos alejados de la distribución principal.

Usando esta información de la distribución se tomó la decisión de no eliminar los posibles datos atípicos de humedad relativa, además de tratarse de valores fácilmente posibles dentro experimento realizado.

Respecto a la humedad del sustrato se realizó el mismo análisis y se tuvo en cuenta que la medición de esta variable se ve afectada también por el cambio de temperatura, sobre todo cuando la temperatura baja durante ciertas horas del día. Haciendo que la medición de humedad del sustrato sea 100% durante muchos instantes del historial de medición.

Figura 49

Diagrama de caja porcentaje de humedad de sustrato grupo de sensores 1 ambiente semi controlado.

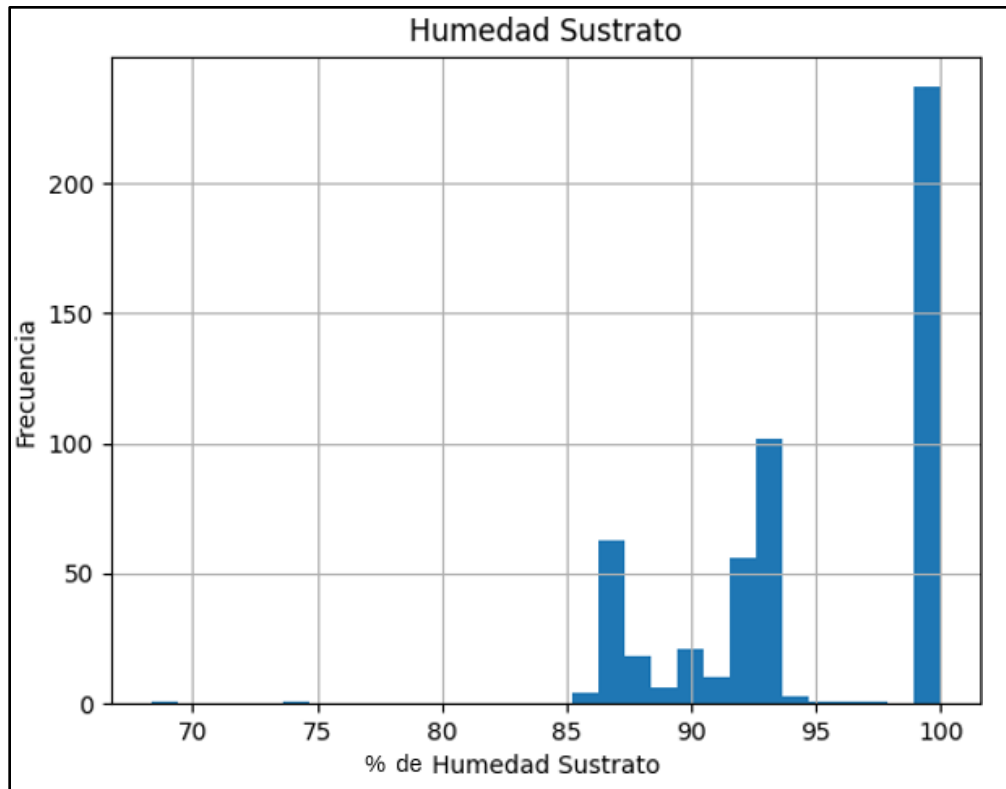


Nota. Diagrama de caja para porcentaje de la variable humedad de sustrato del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

No se evidencian valores atípicos fuera de lo posible en el ambiente en que se encuentran las plantas. Durante el experimento se evidenció que el sustrato utilizado mantiene mucha humedad por lo que los valores altos son apropiados.

Figura 50

Histograma humedad de sustrato grupo de sensores 1 ambiente semi controlado.

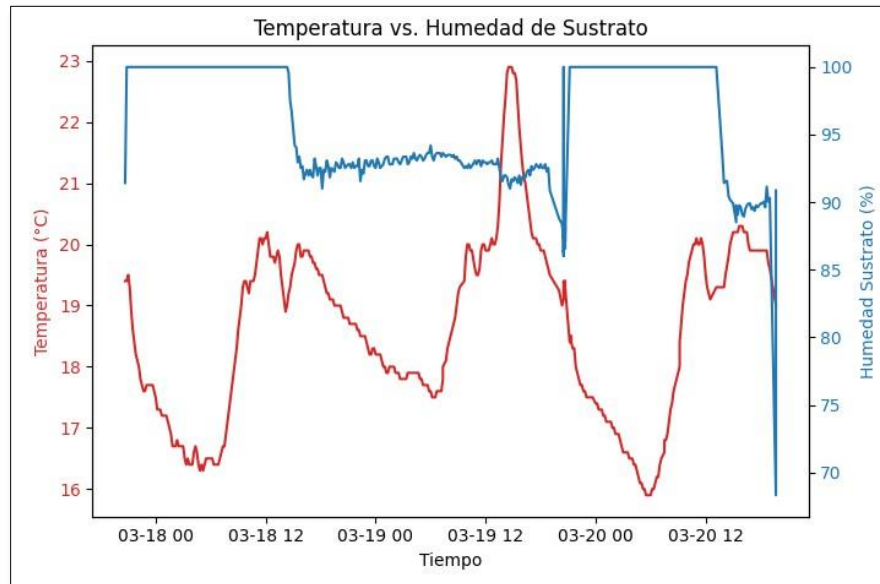


Nota. Histograma para la variable humedad de sustrato del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

En este caso se pueden observar dos estados de la humedad relativa, cuando la temperatura es baja y cuando no, cómo se observa en la siguiente figura.

Figura 51

Temperatura vs Humedad de Sustrato grupo de sensores 1 ambiente semi controlado.

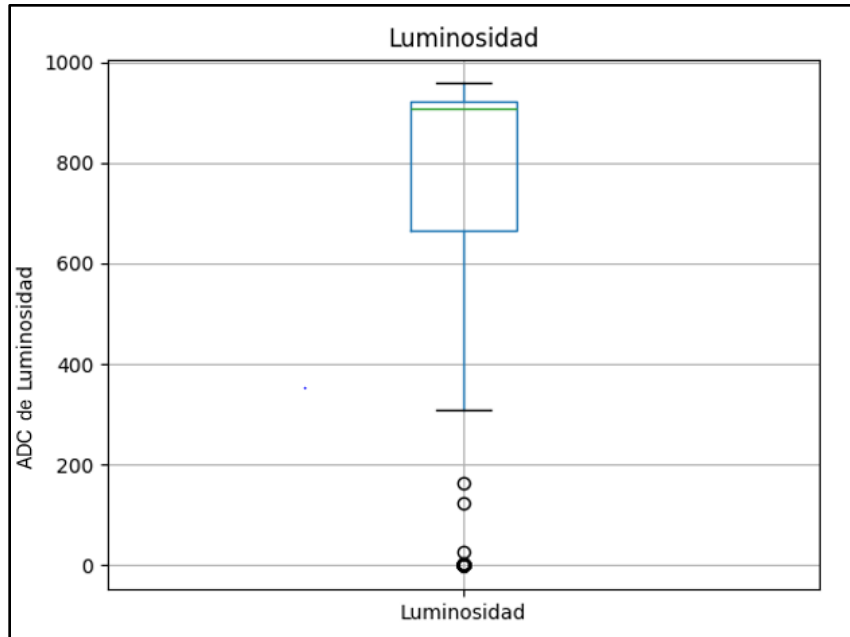


Nota. Temperatura vs humedad de sustrato, se puede ver que durante los puntos de temperatura más baja la humedad del sustrato tendía a ser 100%.

La luminosidad encontrada cuenta con un comportamiento similar a la humedad de sustrato, pero en este caso tiene etapas donde la luminosidad es cero, tratándose de cuando las luces artificiales se encontraban a apagadas siendo esto durante las noches y luego valores de luminosidad aceptables para lo esperado en el experimento.

Figura 52

Diagrama de caja ADC de luminosidad real grupo de sensores 1 ambiente semi controlado.

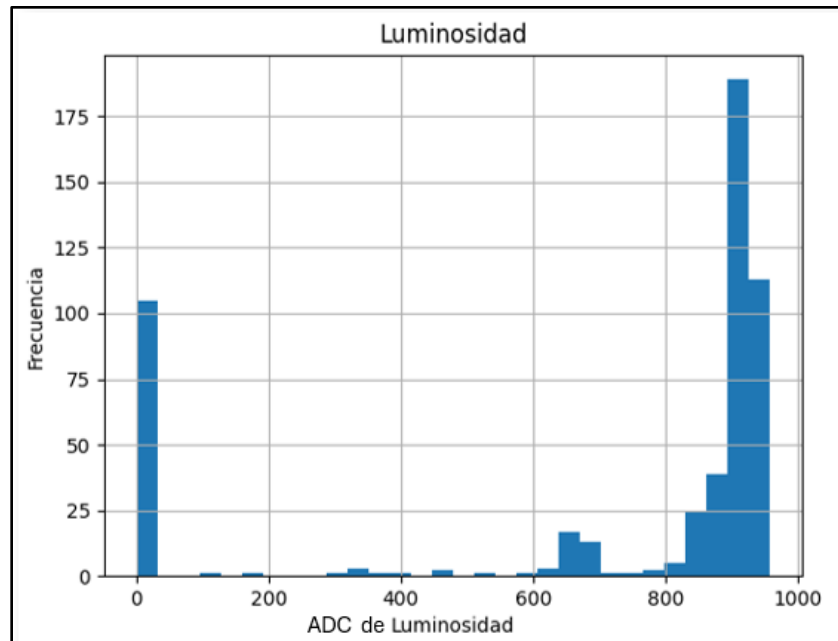


Nota. Diagrama de caja para la variable luminosidad del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

En la figura anterior se pueden ver valores fuera de los bigotes del diagrama, pero estos son los valores que corresponden al momento en que las plantas no tienen luz por lo que no se consideran outliers.

Figura 53

Histograma luminosidad real grupo de sensores 1 ambiente semi controlado.



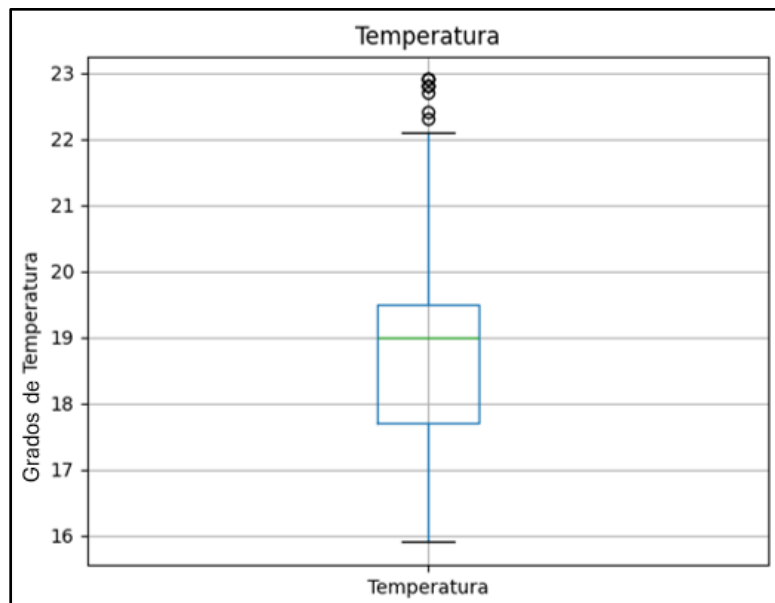
Nota. Histograma para la variable luminosidad del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

Al igual que en la figura 52 (diagrama de caja luminosidad) se evidencian dos grupos correspondientes a los distintos estados de las luces utilizadas en las plantas.

La temperatura encontrada mantiene valores dentro de lo esperable, entre 16 °C y 23 °C, por lo que tampoco se consideró necesario eliminar outliers.

Figura 54

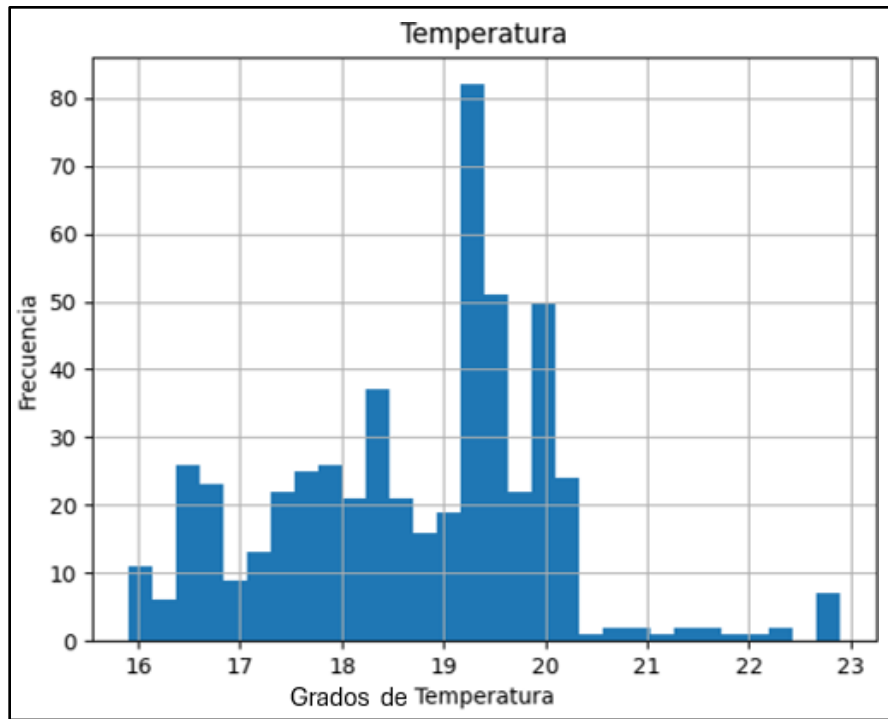
Diagrama de caja temperatura real grupo de sensores 1 ambiente semi controlado.



Nota. Diagrama de caja para la variable temperatura del grupo de sensores 1 del ambiente semi controlado, los datos fuera de los bigotes se encuentran dentro de un rango aceptable de posibles valores de temperatura. Elaboración propia.

Figura 55

Histograma temperatura real grupo de sensores 1 ambiente semi controlado.



Nota. Histograma para la variable temperatura del grupo de sensores 1 del ambiente semi controlado. Elaboración propia.

Respecto a los rayos ultravioleta encontrados no se evidenció un cambio considerable en el estado de esta variable por lo que más adelante se tendrá en cuenta para ser descartada de los modelos generados.

Luego de realizar el anterior análisis del tipo de datos y sus rangos aceptables y analizar las distribuciones en los diferentes gráficos, se constató que los outliers presentes son escasos o son naturales dentro de la distribución. Por lo tanto, se decidió no eliminarlos al considerar que no representan errores de medición y no generan un impacto significativo en los resultados finales del modelo.

2.7.6. Datasets para Modelo SARIMA

Para generar los datos sintéticamente se consideraron las siguientes observaciones que pueden agregar información a los datos nuevos actuando como regresores exógenos. Estos regresores son independientes de los valores recolectados dentro del experimento por lo que no están correlacionados directamente con el término de error del modelo (Gujarati y Porter, 2009, p. 468).

Las variables para generar datos sintéticos y sus posibles regresores adicionales son las siguientes:

- Temperatura: se integra la temperatura mínima y máxima para cada día que se va a generar, de esta manera se ayuda al modelo a captar ciclos de temperatura más amplios que no se pueden observar en los datos reales.
- Humedad de sustrato: se integra como regresor exógeno la temperatura, como se pudo ver en la figura 48 (temperatura vs humedad de sustrato) la temperatura y la humedad de sustrato están relacionadas.
- Humedad relativa, luminosidad y UV: no se considera ningún regresor adicional.

Adicionalmente, se agrega el minuto del día para que el modelo pueda captar los cambios durante el día y generarlos para otros días.

En este caso con SARIMA y ARIMA se buscó reconstruir datos pasados que no fueron medidos, por esta razón, el proceso a llevar a cabo fue hacer *backcast* o predicción hacia atrás para hallar esos valores faltantes.

Una vez comprendido esto se generaron los siguientes datasets para entrenamiento y para predicción representados en las siguientes tablas.

Tabla 23

Estructura dataset para entrenar modelo de datos sintéticos

Timestamp	Temperatura	Temp_high	Temp_low	Minutos_inicio_dia
2025-03-18 00:00:00	17.5	20.0	10.0	0
2025-03-18 04:10:00	16.6	20.0	10.0	250
2025-03-18 10:00:00	19.3	20.0	10.0	600

Nota. El dataset contenía 288 registros entre 18 de marzo de 2025 a las 00:00 hasta el 19 de marzo de 2025 a las 23:50. Los datos de temperatura alta y temperatura baja durante el día (Temp_high y Temp_low) fueron tomados de [accuweather.com](https://www.accuweather.com) para Facatativá, Cundinamarca.

Los datos para predecir se encontraban dentro del dataset con la estructura presentada en la siguiente tabla. Estos datos se encontraban entre el 1 de enero de 2024 hasta el 17 de marzo de 2025, de esta manera se podría completar la serie.

Tabla 24*Estructura dataset para predecir modelo de datos sintéticos*

Timestamp	Temp_high	Temp_low	Minutos_inicio_día
2024-01-01 00:00:00	22.0	12.0	0
2024-09-29 19:50:00	16.0	11.0	1190
2025-03-17 23:50:00	18.0	12.0	1430

Nota. El dataset contenía 63649 registros entre 1 de enero de 2024 a las 00:00 hasta el 17 de marzo de 2025 a las 23:50. Los datos de temperatura alta y temperatura baja durante el día (Temp_high y Temp_low) fueron tomados de accuweather.com para Facatativá, Cundinamarca.

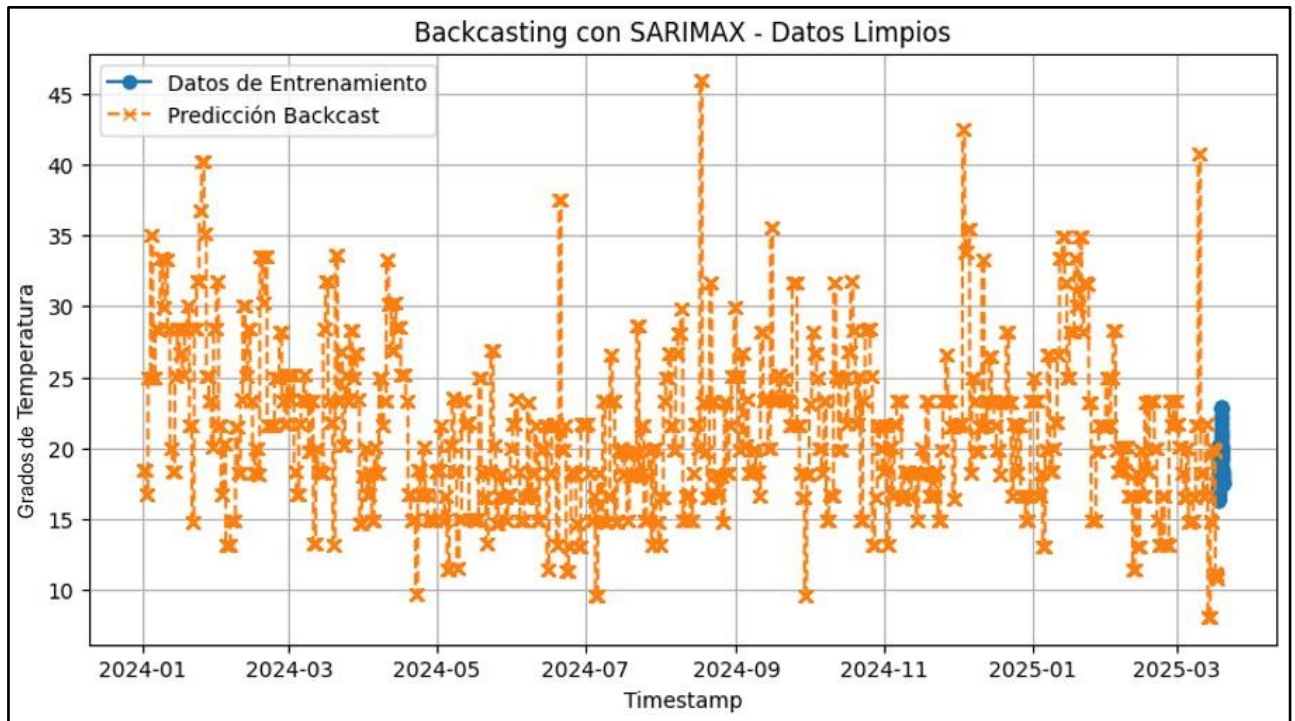
2.7.7. Backcasting de Datos Sintéticos ARIMA

Para los modelos hechos a continuación se usó la librería *statsmodels* 0.14.4 y sus módulos *statsmodels.tsa.statespace.sarimax* (SARIMA o SARIMAX) y *statsmodels.tsa.arima.model* (ARIMA).

Para el primer modelo SARIMAX se usaron los siguientes valores: $order = (1, 0, 1)$ y $seasonal_order = (1, 0, 1, 144)$. Esto significa que el modelo tomó en cuenta el valor anterior de la serie y el error anterior para realizar las predicciones, sin aplicar transformaciones porque los datos ya se consideraban estables. Además, se modeló una componente estacional ya que existía un patrón que se repetía cada 144 unidades de tiempo, incluyendo también los valores y errores del ciclo anterior, pero no se obtuvieron valores aceptables a simple vista. Se usaron como regresores exógenos la temperatura más alta y baja para el día en concreto.

Figura 58

Predicciones SARIMAX con dos días de datos de entrenamiento.

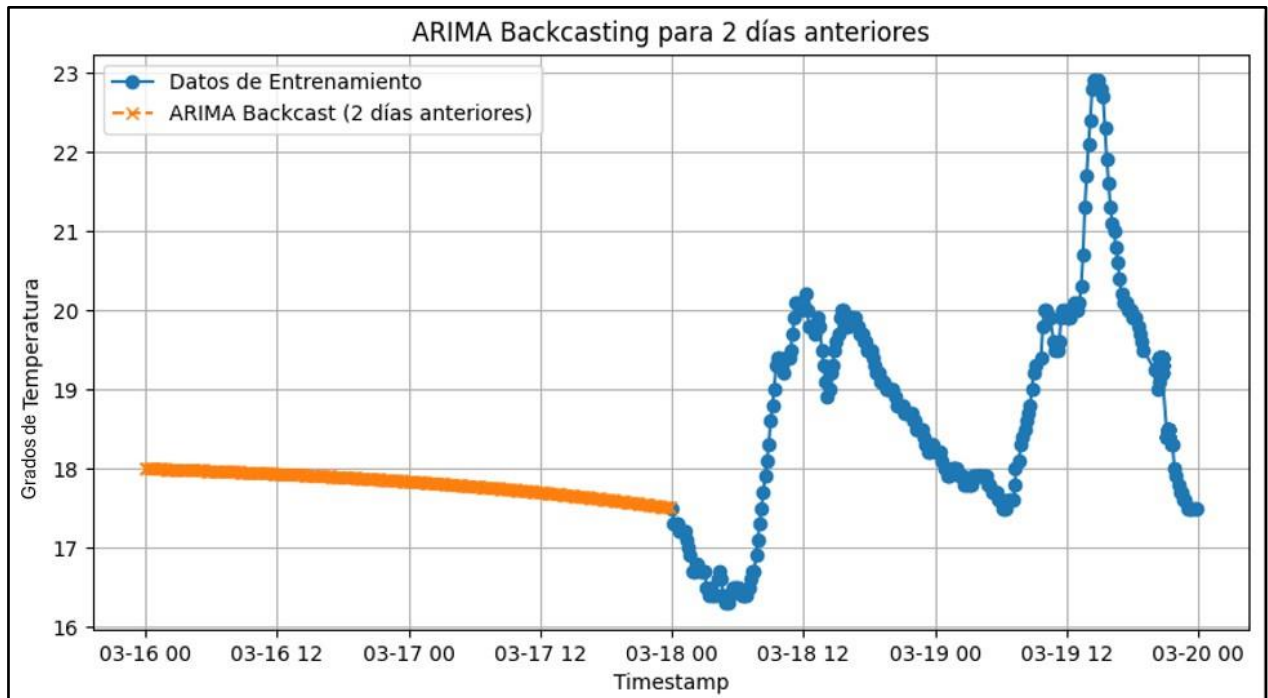


Nota. Se puede observar que los valores generados por el modelo se alejan bastante de los datos originales. Elaboración propia.

Además de SARIMA también se entrenó un modelo ARIMA, sin el componente estacional pero tampoco se obtuvieron resultados aceptables cómo se puede ver en la siguiente figura ya que el modelo no es capaz de captar la estacionalidad de los datos

Figura 59

Predicciones ARIMA para dos días anteriores.



Nota. Se puede observar que el modelo no capta los ciclos diarios de los datos.

Elaboración propia.

Para esta etapa de generación de datos al no obtener los resultados esperados, se optó por usar otro tipo de modelo que se acomodara a la forma de los datos sin intentar captar el componente de ciclos diarios, dada la poca información real disponible.

2.7.8. Backcasting de Datos Sintéticos con red neuronal

Con la intención de capturar la forma no lineal de los datos se pensó en hacer uso de una red neuronal, dada su capacidad para adaptarse a patrones propios de los datos (Varela-Arregocés y Campbells, 2011, p 23). Para esto se usó una red neuronal estructurada

con tres capas ocultas de 64, 32 y 16 neuronas respectivamente usando la función de activación ReLU y una sola neurona de salida para realizar la regresión del dato a predecir.

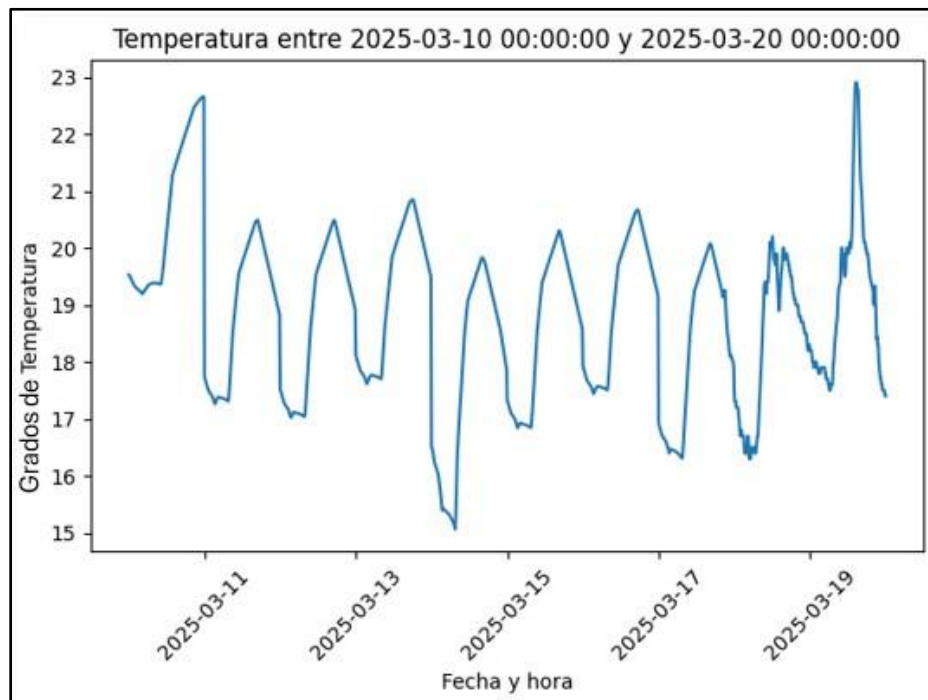
Adicionalmente buscando ajustarse lo máximo posible a la forma de los datos se emplearon 300 épocas lo cual favoreció a un ajuste más fino sobre la información real en comparación con otro modelo de menos épocas no presente en este documento.

Usando la misma red neuronal se hicieron predicciones para las siguientes variables:

- Temperatura: Se hizo uso del mismo dataset presente en la tabla 23 (dataset entrenamiento) para el entrenamiento menos el timestamp. Se obtuvieron los resultados en la siguiente imagen, donde a partir del 17 de marzo de 2025 son datos reales y los demás son generados por el modelo.

Figura 60

Datos sintéticos y reales de grados de temperatura.

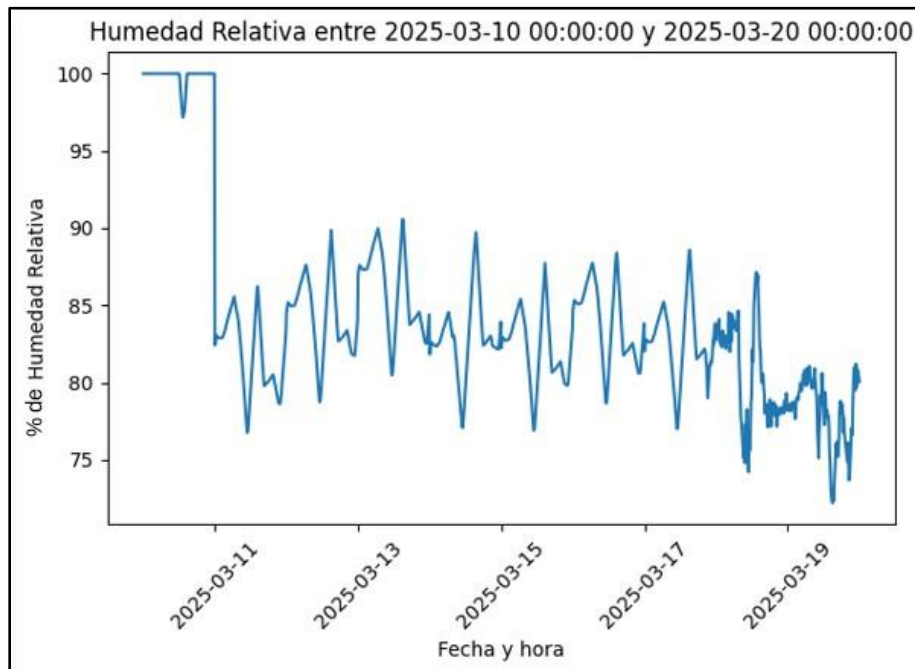


Nota. Datos de temperatura sintéticos y reales entre el 10 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

- Humedad relativa: Para esta variable además del dataset anterior se usó la temperatura sintética y real como otra variable predictora. La temperatura como se vio anteriormente influye en la humedad percibida por los sensores. Obteniendo los siguientes resultados.

Figura 61

Datos sintéticos y reales de porcentaje de humedad relativa.

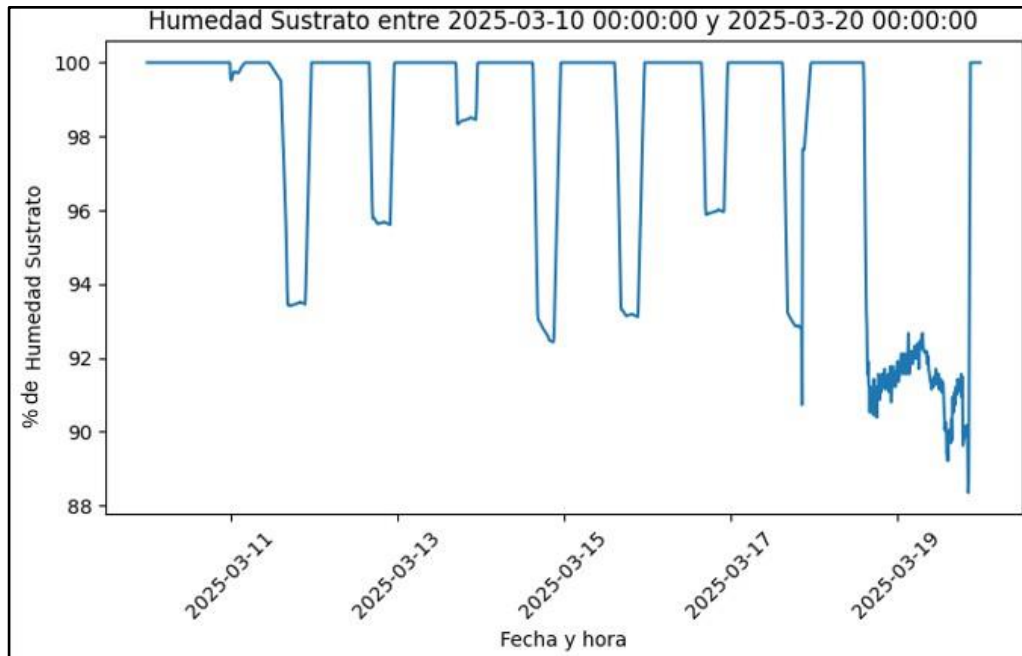


Nota. Datos del porcentaje de humedad relativa sintéticos y reales entre el 10 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

- Humedad de sustrato: Para esta variable además del dataset anterior se usó la temperatura sintética y real como otra variable predictora. Obteniendo los siguientes resultados.

Figura 62

Datos sintéticos y reales de porcentaje de humedad de sustrato.

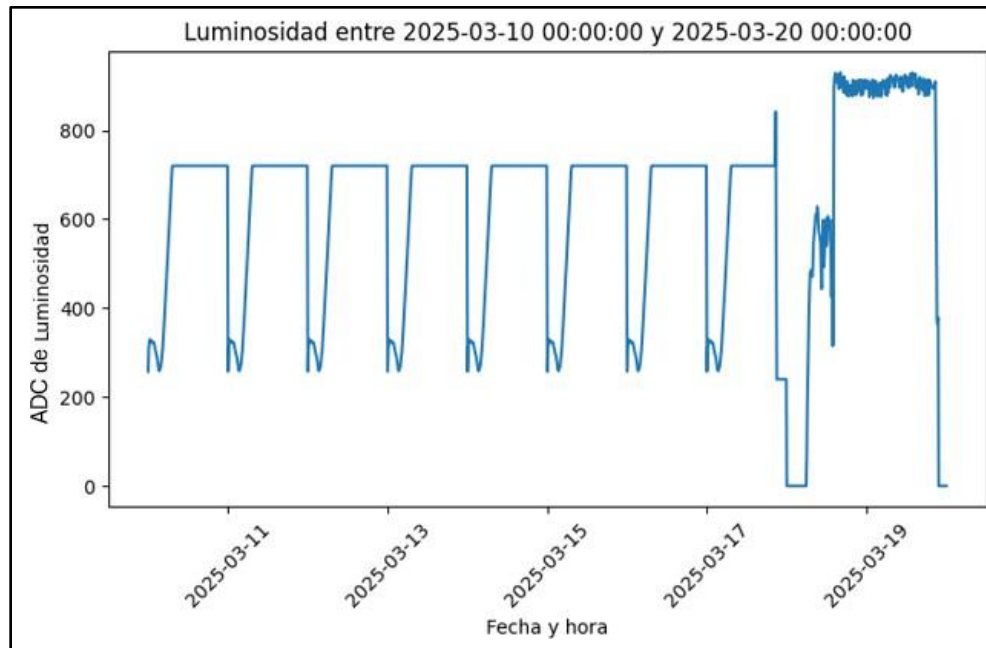


Nota. Datos del porcentaje de humedad de sustrato sintéticos y reales entre el 10 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

- Luminosidad: Para esta variable al no contar con más información que el paso del tiempo durante el día fue complicado encontrar un patrón válido generado por el modelo. En el experimento se usan luces artificiales aparte de la luz solar que puede llegar a entrar, pero información de cuando se encendían y cuando no, no se tiene por lo que es difícil modelar lo que sucede con esta variable fuera el ciclo día y noche que sí pudo ser captado por el modelo.

Figura 63

Datos sintéticos y reales de luminosidad.



Nota. Datos de luminosidad sintéticos y reales entre el 10 de marzo de 2025 y el 20 de marzo de 2025. Elaboración propia.

Este mismo proceso se aplicó para los diferentes grupos de sensores en cada uno de los ambientes que se tenían presentes.

2.7.9. Variable Dependiente (altura de la planta)

Respecto a esta variable se contaba con la altura de la planta, pero con mediciones ocasionales. Por esta razón, fue necesario reconstruir la altura usando esos puntos en el tiempo cómo referencia. Para esto se contaba con información en la tabla a continuación.

Tabla 25

Registros de altura de las plantas en centímetros.

Altura (cm) ambiente semi controlado			
Fecha	Planta 1	Planta 2	Planta 3

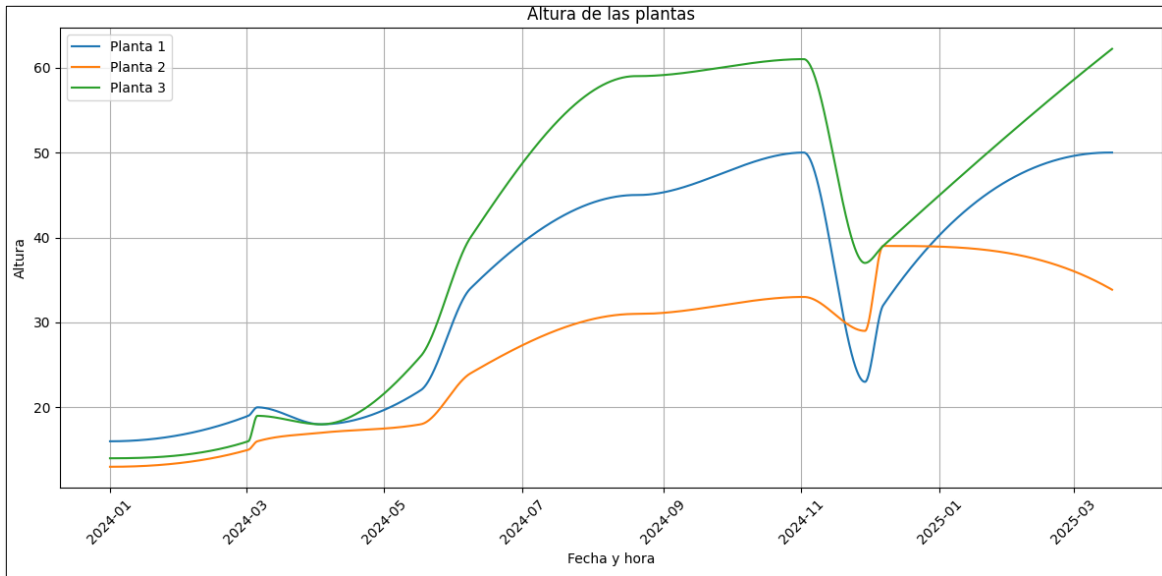
2/3/2024	19	15	16
6/3/2024	20	16	19
3/4/2024	18	17	18
17/05/2024	22	18	26
6/8/2024	34	24	40
20/8/2024	45	31	59
21/8/2024	45	31	59
29/11/2024	23	29	37
7/12/2024	32	39	39
02/07/2025	57	34	74

Nota. Registros de altura de las plantas del ambiente semi controlado en diferentes fechas.

Elaboración propia.

Con la información de la tabla anterior se interpolaron datos para completar toda la serie temporal, al igual que en las variables ambientales.

Figura 64
Altura de las plantas



Nota. Acá se representa la altura de las plantas, el 2 de noviembre hay una caída en la altura ya que en esta fecha se hizo poda de formación. Elaboración propia.

Las plantas sufrieron diferentes podas a lo largo del tiempo, pero la más significativa sucedió el día 2 de noviembre de 2024. Esta poda redujo la altura de las plantas, lo que afecta la tendencia de la altura de las plantas, por esta razón, en el modelo se consideró hacer el entrenamiento hasta la fecha de esta poda.

En el capítulo “Ambientes y Plantas” se explica que inicialmente se planteó un prototipo computacional que recopilaba datos de tres ambientes diferentes, cada uno con sus propias características y tipos de plantas. La idea era comparar estos ambientes para determinar cuáles ofrecían las condiciones más favorables para el cultivo de arándanos Biloxi. Sin embargo, debido a la disponibilidad de datos, se decidió trabajar con uno solo de esos ambientes para construir los modelos y, así, evaluar cuál de ellos explicaba mejor la altura de las plantas considerando el entorno.

Además, se promedió la altura de las tres plantas, ya que todas crecieron en las mismas condiciones ambientales. Así se pudo simplificar el modelo matemático a uno por ambiente y de esta manera eventualmente poder compararlo con otros ambientes.

Adicionalmente, se hizo el mismo proceso de promediar para los dos grupos de sensores.

Para finalizar este capítulo, se generó un dataset que integra tanto las variables ambientales y la altura de las plantas para cada registro temporal.

Tabla 26

Estructura dataset para entrenamiento de modelos

timesta	temp_hi	temp_l	hum_r	min_siem	hum_su	lum	temp	altur
mp	gh	ow	el	bra	str			a
2024-								
01-01								
00:00:0	22.0	12.0	91.029	87540	100.0	257.5	17.6	14.3
0						63	81	33
2024-								
05-19								
07:30:0	20.0	10.0	84.366	288150	99.87	720.1	17.3	22.5
0						95	11	89

Nota. Dataset para el entrenamiento de los modelos, se encuentran las variables generadas sintéticamente y luego promediadas para el ambiente semi controlado. Elaboración propia

2.8. Análisis de Modelos

En esta etapa se exponen los modelos que más tuvieron relevancia durante la investigación realizada, en diferentes tablas con el fin de poder seleccionar al menos dos de estos que mejor se acomoden a lo que se quiere llegar al final, que es tener una descripción y comprensión del crecimiento de las plantas de arándanos en su variedad biloxi.

2.8.1. Modelo Aditivo GAM

Los **Modelos Aditivos Generalizados (GAM, por sus siglas en inglés)** son una extensión de los modelos lineales generalizados (GLM) que permiten modelar relaciones no lineales entre las variables predictoras y la variable respuesta utilizando funciones suaves en lugar de coeficientes fijos (Hastie & Tibshirani, 1990). En términos generales

$$y_i = \mathbf{B}_0 + \sum_{j=1}^p \left(\sum_{k=1}^{K_j} \mathbf{B}_{jk} b_{jk}(x_{ij}) \right)$$

Donde:

- y_i = Variable objetivo para el registro i
- \mathbf{B}_0 = Término de intercepción
- p = Número de variables predictoras
- K_j = Número de funciones base asociados al predictor j
- \mathbf{B}_{jk} = Coeficiente del modelo para función b_k del predictor j
- $b_{jk}(x_{ij})$ = Funciones base evaluadas en el valor X_{ij}

Este enfoque permite capturar relaciones no lineales complejas sin necesidad de especificar una forma funcional rígida, y se emplea comúnmente en contextos donde se

requiere mayor flexibilidad que en la regresión lineal tradicional, pero con control sobre el sobreajuste mediante penalizaciones (Wood, 2017).

2.8.2. Modelo GAM para el Crecimiento de las Plantas

Aplicación de los GAM en el Crecimiento de las Plantas

El crecimiento de las plantas es un proceso complejo influenciado por múltiples factores ambientales, como la temperatura, la luz solar, la disponibilidad de agua y los nutrientes del suelo (Poorter et al., 2012). Los modelos GAM han sido utilizados en estudios ecológicos y agrícolas para capturar relaciones no lineales entre estas variables y la tasa de crecimiento de las plantas (Dormann et al., 2008).

- **Modelado de relaciones no lineales:** El crecimiento de las plantas no sigue siempre una tendencia lineal, ya que puede haber períodos de aceleración y desaceleración debido a cambios en el ambiente (Guisan et al., 2002).
- **Identificación de factores críticos:** Los GAM permiten detectar qué variables ambientales tienen un mayor impacto en el crecimiento de las plantas y en qué rangos específicos (Wood, 2017).
- **Predicciones más precisas:** Debido a su flexibilidad, los GAM pueden mejorar la precisión en la estimación del crecimiento futuro de las plantas bajo diferentes condiciones climáticas (Dormann et al., 2008).
- **Los Modelos Lineales (ML):** representan una de las herramientas más utilizadas para estudiar la relación entre variables predictoras y una variable dependiente en la investigación científica. Su simplicidad y facilidad de interpretación los hacen

ideales para evaluar el efecto de factores ambientales como la luz, la temperatura y los nutrientes en el crecimiento de las plantas (Montgomery et al., 2012).

2.8.3. Modelos no Lineales para Curva Sigmoide

El crecimiento de las plantas es un proceso complejo influenciado por múltiples factores ambientales y fisiológicos. A diferencia de los modelos lineales, que asumen una relación constante entre las variables, los modelos no lineales permiten capturar patrones más realistas, como el crecimiento acelerado seguido de una fase de estabilización (Larcher, 2003). Uno de los modelos más utilizados en este contexto es la curva en forma de S o sigmoide, que describe cómo las plantas crecen rápidamente en una etapa temprana, alcanzan una tasa máxima de crecimiento y luego se estabilizan a medida que se acercan a su tamaño máximo (Yin et al., 2003).

Los modelos sigmoides representan el crecimiento vegetal mediante una ecuación de la forma:

$$y = \frac{K}{1 + e^{-r(t-t_0)}}$$

Donde:

- **Y** es el tamaño de la planta en un momento dado,
- **K** es el crecimiento máximo alcanzado,
- **r** es la tasa de crecimiento,
- **t** es el tiempo,
- **t₀** es el punto en el que la tasa de crecimiento es máxima,
- **e** es la base del logaritmo natural.

Este modelo refleja cómo las plántulas inicialmente crecen lentamente, luego alcanzan una fase de rápido crecimiento exponencial y finalmente se estabilizan debido a factores limitantes como nutrientes, espacio y competencia (Hunt, 1990).

2.8.4. Beneficios del Modelo Sigmoide en la Descripción del Crecimiento Vegetal

- **Captura la dinámica real del crecimiento:** A diferencia de los modelos lineales, la curva sigmoide representa de manera más precisa la fase inicial lenta, el rápido crecimiento intermedio y la estabilización final (Polo, V. 2013).
- **Permite predecir el tamaño final de la planta:** Al ajustar datos experimentales a una curva sigmoide, se puede estimar el tamaño máximo que alcanzará una planta bajo determinadas condiciones ambientales (Yin et al., 2003).
- **Facilita la identificación de factores limitantes:** La fase de estabilización del modelo puede ayudar a determinar qué factores ambientales restringen el crecimiento, como la disponibilidad de agua o nutrientes (Larcher, 2003).

2.8.5. Análisis de Series de Tiempo

El análisis de series de tiempo es una herramienta estadística utilizada para estudiar patrones y tendencias en datos recolectados a lo largo del tiempo (Chatfield, 2003). En la biología vegetal, este enfoque permite modelar y predecir el crecimiento de las plantas considerando factores como temperatura, humedad y disponibilidad de nutrientes. A diferencia de los modelos estáticos, el análisis de series de tiempo capta la naturaleza dinámica del crecimiento vegetal y ayuda a comprender cómo las condiciones ambientales afectan su desarrollo (Zuur et al., 2007).

2.8.6. Series de Tiempo el Crecimiento de las Plantas

Las series de tiempo son secuencias de datos recopilados en intervalos regulares, como la altura de una planta medida diariamente o el índice de área foliar registrado semanalmente, pudiendo resaltar las siguientes características

- **Tendencias:** patrones de crecimiento a largo plazo, como un aumento constante en la biomasa.
- **Estacionalidad:** variaciones periódicas en el crecimiento debidas a cambios climáticos o estacionales (Box et al., 2015).
- **Componentes aleatorios:** fluctuaciones inesperadas en el crecimiento debido a factores externos como plagas o enfermedades (Wei, 2006).

Para modelar estas características, se utilizan métodos como:

- **Modelos Autorregresivos Integrados de Media Móvil (ARIMA):** Son útiles para capturar tendencias y fluctuaciones en el crecimiento vegetal mediante la combinación de términos autorregresivos, diferencias y promedios móviles (Box et al., 2015).
- **Modelos de Suavizamiento Exponencial:** Permiten predecir el crecimiento a corto plazo dando más peso a las observaciones recientes (Hyndman & Athanasopoulos, 2018).
- **Modelos de Procesos Estocásticos:** Capturan la variabilidad aleatoria en el crecimiento de las plantas y su respuesta a condiciones ambientales cambiantes (Zuur et al., 2007).

El análisis de series de tiempo ha sido ampliamente utilizado para estudiar el crecimiento de cultivos y especies forestales. Algunos ejemplos de su aplicación incluyen:

- **Predicción del rendimiento agrícola:** Modelos ARIMA han sido utilizados para predecir la producción de cultivos en función de factores climáticos históricos (Box et al., 2015).
- **Monitoreo del crecimiento de plántulas en invernaderos:** Series de tiempo ayudan a optimizar el uso de fertilizantes y agua al identificar patrones de respuesta en el crecimiento vegetal (Wei, 2006).
- **Evaluación del impacto del cambio climático:** Permite analizar cómo variables climáticas afectan el crecimiento de especies vegetales a lo largo del tiempo (Hyndman & Athanasopoulos, 2018).

2.8.7. Redes Neuronales Recurrentes RNN

Las redes neuronales recurrentes (RNN, por sus siglas en inglés) son un tipo de modelo de inteligencia artificial diseñado para procesar datos secuenciales. A diferencia de las redes neuronales tradicionales, que analizan cada entrada de manera independiente, las RNN tienen la capacidad de retener información previa y utilizarla en la predicción de datos futuros (Hochreiter & Schmidhuber, 1997). Esta característica las hace especialmente útiles para modelar procesos biológicos dinámicos, como el crecimiento de las plantas, donde las condiciones ambientales y los factores fisiológicos cambian con el tiempo (LeCun et al., 2015).

Las RNN son un tipo de red neuronal artificial que se caracteriza por tener conexiones cíclicas, lo que les permite recordar información de entradas anteriores y aprender patrones secuenciales en los datos (Goodfellow et al., 2016). A diferencia de las

redes neuronales feedforward, que procesan la información en una única dirección, las RNN pueden retener estados previos y utilizarlos para predecir valores futuros.

Matemáticamente, la activación de una RNN se define como:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

Donde:

- h_t es el estado oculto en el tiempo t ,
- h_{t-1} es el estado oculto del tiempo anterior,
- x_t es la entrada en el tiempo t ,
- W_h y W_x son matrices de pesos,
- b es un sesgo,
- f es una función de activación no lineal.

Las RNN tradicionales pueden presentar el problema de **gradientes desaparecidos** o **explosivos**, lo que dificulta su entrenamiento en secuencias largas. Para solucionar esto, se han desarrollado variantes como las **Long Short-Term Memory (LSTM)** y las **Gated Recurrent Units (GRU)**, que permiten modelar dependencias a largo plazo de manera más eficiente (Hochreiter & Schmidhuber, 1997).

2.8.8. Funcionamiento de las Redes Neuronales Recurrentes

Las RNN funcionan mediante un proceso iterativo en el que la información de la secuencia anterior influye en la predicción del siguiente paso. Su estructura incluye:

1. **Capa de entrada:** Recibe los datos secuenciales, como mediciones del crecimiento de una planta en diferentes días.
2. **Capas recurrentes:** Procesan los datos secuenciales y almacenan información de estados previos para capturar dependencias temporales.
3. **Capa de salida:** Genera la predicción final, como la estimación del crecimiento futuro de la planta.

Gracias a su capacidad para modelar dependencias temporales, las RNN son especialmente útiles en problemas donde la información pasada influye en eventos futuros, como la dinámica de crecimiento de organismos vivos (Goodfellow et al., 2016).

2.8.9. Aplicación Redes Neuronales Recurrentes en el Crecimiento de las Plantas

El crecimiento de una planta está influenciado por factores como temperatura, humedad, radiación solar y disponibilidad de nutrientes. Las RNN permiten analizar grandes volúmenes de datos ambientales y fisiológicos para predecir patrones de crecimiento con mayor precisión que los modelos tradicionales (LeCun et al., 2015).

2.8.10. ¿Qué es Normalización de Datos?

La normalización es una técnica de escalado de datos que ajusta los valores de las variables a un rango específico, generalmente entre 0 y 1 o con una media de 0 y desviación estándar de 1. Esto evita que variables con diferentes escalas afecten negativamente el rendimiento de los modelos de predicción (Han et al., 2011).

Algunos de los métodos más utilizados incluyen:

- Min-Max Scaling

- Z-Score Scaling (Estandarización)
- Normalización Decimal (Decimal Scaling)
- Transformación Logarítmica

2.8.11. Importancia de la Normalización en Modelos Predictivos del Crecimiento de Plantas

Los modelos como las redes neuronales recurrentes y los modelos de series de tiempo son altamente sensibles a la escala de los datos. Si una variable tiene valores mucho mayores que otra, el modelo puede darle mayor peso, afectando la precisión de las predicciones (LeCun et al., 2015).

En redes neuronales recurrentes (RNN):

- Evita problemas de gradientes explosivos o desaparecidos.
- Asegura que todas las variables tengan la misma relevancia en la predicción del crecimiento de las plantas.

En modelos de análisis de series de tiempo:

- Mejora la estabilidad numérica y reduce la varianza en predicciones climáticas y agrícolas.
- Facilita la detección de patrones en variables como temperatura, humedad y tasas de crecimiento.

En modelos estadísticos tradicionales:

- Mejora la interpretabilidad de los coeficientes en modelos GAM, lineales y no lineales aplicados al crecimiento de las plantas.

Tabla 27
Tabla de características de modelos

Característica	Modelos No			
	Modelos GAM	Modelos Lineales	Lineales Curvas S (Richards/Logística)	Modelo de von Bertalanffy
Tipo de relación que modela	No lineal, aditiva	Lineal	No lineal (curva S)	No lineal (relacionado a biomasa/longitud)
Flexibilidad	Alta	Baja	Moderada	Moderada
Complejidad	Moderada	Baja	Moderada	Moderada
Interpretabilidad	Moderada (efectos individuales de variables)	Alta (coeficientes)	Moderada (parámetros de crecimiento)	Moderada (parámetros de crecimiento)
Adecuado para...	Explorar relaciones no lineales complejas, modelar efectos de múltiples variables	Obtener una primera aproximación, comparar promedios entre grupos	Modelar crecimiento con forma de S, estimar parámetros de crecimiento	Modelar crecimiento en términos de biomasa o longitud
Aporte a la investigación de arándanos	Modelar como unas variables influyen en el crecimiento	Comparar la altura promedio entre ambientes	Modelar cómo el tamaño aumenta con el tiempo en cada ambiente	Estimar la tasa de crecimiento y el tamaño máximo (en biomasa) en cada ambiente

o de las
plantas

Nota. Autoría propia.

Se comparan los diferentes modelos según sus características y su evaluación en aspectos clave que determinan su relevancia para el estudio. Entre estos aspectos se incluyen la complejidad, la flexibilidad y, de manera especial, su aporte a la investigación sobre arándanos. Cada modelo presenta ventajas y desventajas, lo que permite tomar una decisión fundamentada en las características extraídas de la investigación realizada.

2.9. Desarrollo, Ajuste y análisis de los resultados obtenidos por los modelos

2.9.1. GAM

Para el desarrollo del modelo GAM se utilizó como lenguaje de programación Python, con ayuda de diferentes librerías como pandas y numpy para el manejo de datos y arreglo de datos, PyGam para el manejo del modelo GAM esta librería ofrece diferentes variantes GAM, por eso se emplea el módulo de LinearGam (modelo de regresión aditiva lineal), junto a dos parámetros que son importantes para el desarrollo S (Función para variables numéricas continuas (Splines)) y F (Función para variables categóricas (Factoriales)).

Por otro lado, con ayuda de la librería de Sklearn y su modulo Model_selection se divide la cantidad de datos en entrenamiento y prueba con el módulo de train_test_split, para las métricas de evaluación se usó el módulo de metrics usando las siguientes métricas r2_score, means_squared_error, las gráficas con ayuda de matplotlib y por último Joblib para el cargue y guardado de modelos, el cual se en archivo .PLK ya que es un modelo no paramétrico no puede ser expresado en una ecuación explícita.

Estructura del modelo

```
X = df[['temp_high', 'temp_low', 'hum_rel', 'min_siembra', 'hum_sustr', 'lum', 'temp',
'poda_engrosamiento']]
y = df['altura']
```

División en entrenamiento y prueba (80% train, 20% test)

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42) # random_state para garantiza que los resultados sean reproducibles.
```

```
gam = LinearGAM(
    s(0, n_splines=10) +
    s(1, n_splines=10) +
    s(2, n_splines=10) +
    s(3, n_splines=10) +
    s(4, n_splines=10) +
    s(5, n_splines=10) +
    s(6, n_splines=10) +
    f(7)
).fit(X_train, y_train)
```

Predicciones sobre conjunto de prueba

```
y_pred = gam.predict(X_test)
```

Métricas de evaluación

```
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
```

```
print(f"R2 (Porcentaje explicado por el modelo): {r2*100:.2f}%")
```

```
print(f"RMSE (Error cuadrático medio): {rmse:.4f}")
```

Definición de variables predictoras y objetivo

Como se muestra, las variables predictoras corresponden a los datos recolectados por los sensores, tales como temperatura, humedad, luminosidad, tiempo de siembra, poda, entre otros. Estas variables se utilizan para analizar su influencia sobre la variable objetivo, que en este caso es la altura de la planta.

Tabla 28
Descripciones de variables predictoras y categóricas

Spliner	Variable	Variable real
S(0)	temp_high	Temperatura alta
S(1)	temp_low	Temperatura mínima
S (2)	hum_rel	Humedad relativa en aire
S (3)	min_siembra	Tiempo transcurrido desde la siembra
S (4)	hum_sustr	Humedad de sustrato
S (5)	lum	Luminosidad
S (6)	temp	Temperatura media
F (7)	poda_engrosamiento	Poda engrosamiento (Variable categorica)

Nota. Elaboración propia.

En la tabla 28 se muestran los spliners y que variable pertenece a estas, de la misma forma se muestra el significado que representa cada una de estas en el ambiente

Entrenamiento

El conjunto de datos se divide en dos partes: el 80% se utiliza para el entrenamiento del modelo y el 20% restante para la prueba, empleando un valor de random_state=42 para garantizar la reproducibilidad de los resultados.

Para el entrenamiento se utiliza el modelo LinearGAM, el cual permite sumar efectos no lineales de cada variable. Este modelo emplea distintos tipos de funciones según la naturaleza de los datos: s(i) para suavizar variables numéricas, f(i) para tratar variables categóricas y fit() para ajustar el modelo a los datos de entrenamiento.

Una vez entrenado, el modelo se guarda utilizando el método. dump() y posteriormente se carga con .load() mediante la librería joblib, lo que permite realizar

predicciones sobre el conjunto de prueba y llevar a cabo el análisis correspondiente.

Métricas de pruebas para el Modelo GAM

```
# Predicciones sobre conjunto de prueba
y_pred = gam.predict(X_test)

# Métricas de evaluación
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print(f"R² (Porcentaje explicado por el modelo): {r2*100:.2f}%")
print(f"RMSE (Error cuadrático medio): {rmse:.4f}")
```

Las métricas de pruebas que se emplean con ayuda de la librería Sklearn se usa el R² (Coeficiente de determinación) el cual mide que tan bien las predicciones explican la variabilidad de los datos reales y el RMSE el cual en cuanto menor sea mejor es, ya que mide cuanto error hay en promedio entre las predicciones y los valores reales.

2.9.2. Análisis de Resultados GAM

Resultados obtenidos en el modelo

Tabla 29

Resultado de métricas

R² (Porcentaje explicado por el modelo):	97.83%
RMSE (Error cuadrático medio):	1.8589

Nota. Elaboración propia.

En la tabla 26 se muestra el resultado de las métricas de análisis obtenidas las cuales permiten entender que el modelo da un buen resultado al ajustarse y predecir los valores basados en la línea de curva central

$$\text{altura} = \mathbf{s(0)} + \mathbf{s(1)} + \mathbf{s(2)} + \mathbf{s(3)} + \mathbf{s(4)} + \mathbf{s(5)} + \mathbf{s(6)} + \mathbf{f(7)} + \mathbf{intercepto}$$

La ecuación explica la estructura del modelo GAM, Lo que hace es sumar el efecto de cada función suavizada spline $S(i)$ o variable predictora, la función lineal de la variable categórica $F(7)$ la cual corresponde a la poda de engrosamiento expresada como una variable categórica y el intercepto (valor base).

Tabla 30

Resultado de la regresión con LinearGAM (Generalized Additive Model)

Distribution:	NormalDist	Effective DoF:	56.7993
Link Function:	IdentityLink	Log Likelihood:	118527.1906
Number of Samples:	50934	AIC:	237169.9798
AICc:			237170.1134

Nota. Elaboración propia.

En la tabla 30 la distribución describe como el modelo asume una distribución normal de los errores, el linkfunction es la función que conecta los predictores con la variable objetivo, el número de muestras que se usaron, AIC/AICc Criterios de Akiike usados para evaluar la calidad del modelo. Menor AIC es mejor.

Tabla 31

Aporte de las variables al modelo

Feature Function	Lambda	Rank	EDoF	P > x	Sig. Code
s(0)	[0.6]	10	8.8	1.11e-16	***
s(1)	[0.6]	10	8.5	1.11e-16	***
s(2)	[0.6]	10	7.6	1.11e-16	***
s(3)	[0.6]	10	8.5	1.11e-16	***
s(4)	[0.6]	10	7.8	6.66e-16	***
s(5)	[0.6]	10	7.6	7.35e-01	

s(6)	[0.6]	10	7.0	3.43e-04	***
f(7)	[0.6]	2	1.0	1.11e-16	***
Significance codes:				C	'***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1

Nota. Autoría propia.

En la tabla 31 se describe los criterios del modelo como lo son: S(I) indica la función suavizada del spline para modelar el efecto no lineal de cada variable del datasets, F(7) es una función lineal de variable categórica, Lambda es el valor de penalización que regula la suavidad del spline (mayor lambda = menos flexibilidad), Rank / EDoF (Grados de libertad efectivos) es el que indica la complejidad del spline, entre mayor EDoF quiere decir que es más complejo, P>x es el valor de P que indica la significancia estadística, el cual si es muy bajo quiere decir que la variable no contribuye significativamente al modelo y por último el Sig Code lo cual por medio de asteriscos muestra la importancia estadística, donde con “***” es muy significativa, “.” Marginal y sin asteriscos no es significativa.

Análisis de los efectos parciales de cada una de las variables objetivos usadas en el modelo

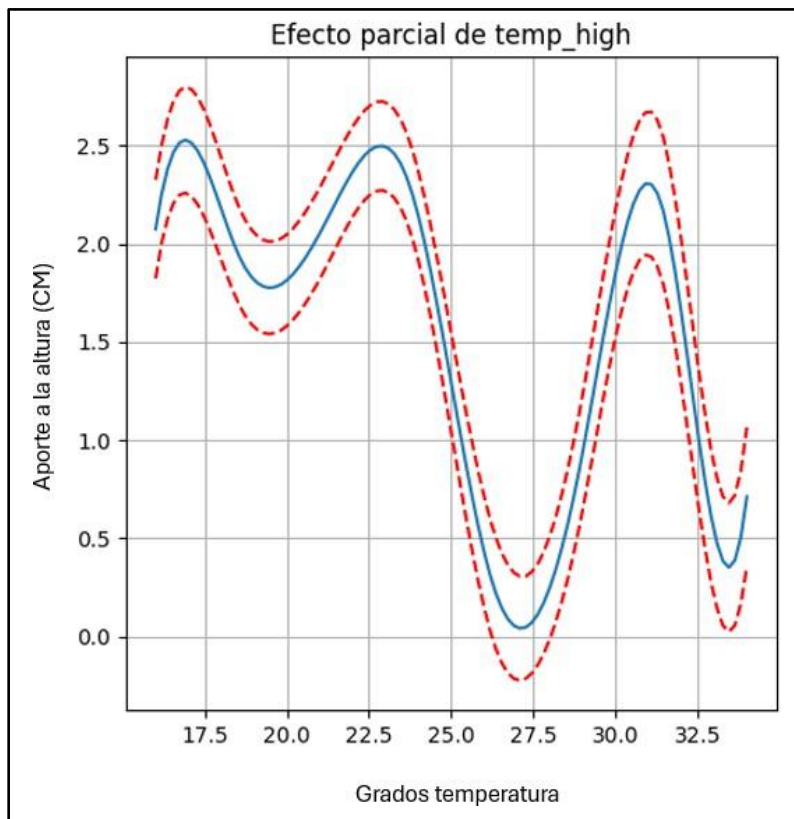
En este apartado se presentan graficas donde la línea azul muestra la representación de la variable y las líneas rojas son las que representan el modelo GAM, en donde se usan términos suaves para la mayoría de las variables continuas y un término factorial para la variable de poda de engrosamiento, lo que permitió tener un resultado de ajuste de R² de (97.83%) lo que indica que el modelo “describe o explica” casi toda la variación de la altura, teniendo presente que los datos generados de cierta forma se consideran a los ideales tiene una relevancia en el resultado de ajuste.

Como se trata de un modelo lineal, cada variable tiene un efecto proporcional sobre el crecimiento de la planta. Los valores con efecto menor a cero van a restar al resultado final, es decir, van a tener un impacto negativo en el crecimiento. En cambio, entre más alto sea el valor del efecto (positivo), mayor será su contribución al crecimiento. De esta forma, el modelo suma o resta según el peso que tenga cada variable.

En este caso, las únicas variables reales medidas en su completitud son la temperatura máxima y la temperatura mínima sobre Facatativá, usándose como fuente de información verídica al modelo. Las demás variables utilizadas en el modelo fueron generadas de forma sintética.

Figura 65

Grafica de efecto parcial de temp_high $S(0)$ vs altura



Nota. Autoría propia.

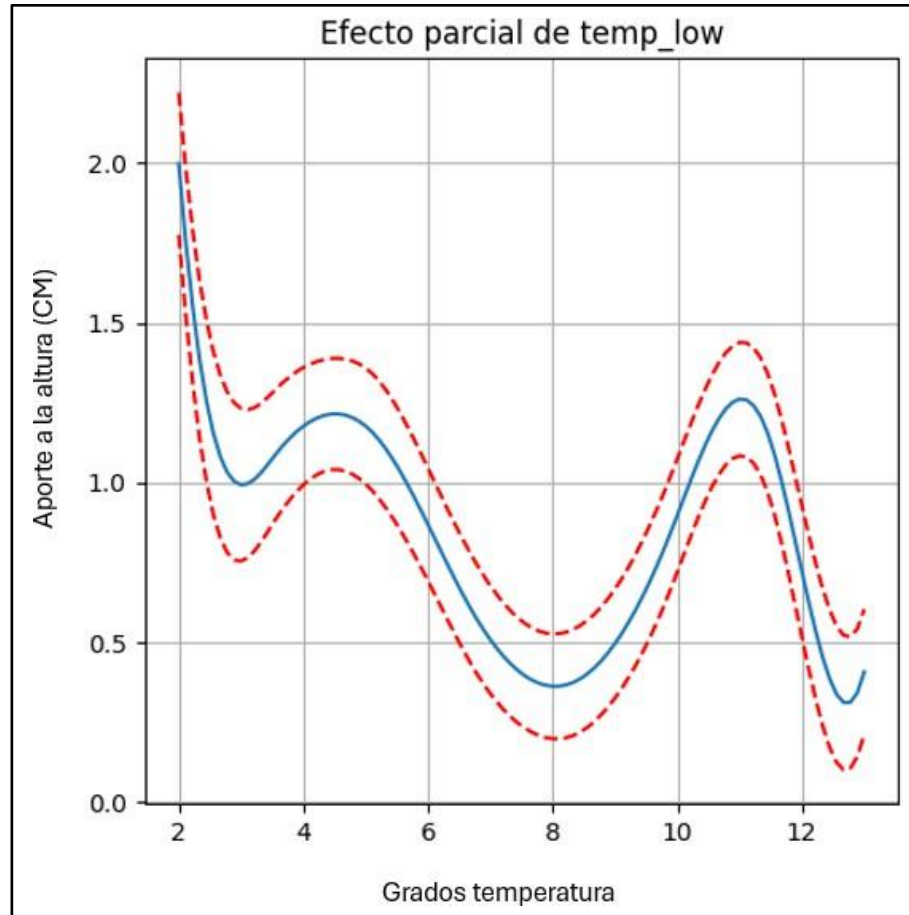
En la figura 65 se muestra la gráfica de la relación de la altura con la variable temperatura máxima que se presenta, en donde la curva presenta oscilaciones o picos y valles lo que indica que la relación entre la temperatura máxima y la altura no es lineal.

Se puede deducir que los rangos donde más se aporta a la altura de la planta podrían estar entre 18 y 23 grados Celsius para que la planta pueda crecer y favorecer a un mayor crecimiento. Por otro lado, en el modelo GAM se emplean la cantidad de splines (10) permite capturar los cambios sutiles de la variable, lo que se traduce como que $p < 0.001$ denotando que el efecto se acerca al real y es relevante dentro del dataset.

Para finalizar basado en lo que se muestra en la gráfica se puede tener presente las siguientes decisiones, poder tener un ajuste dentro de los cultivos para mantener la temperatura máxima en un rango, esto con ayuda de actuadores como ventiladores que permitan disipar el calor en días en donde este sea alto o medidas de simulación de sombreado.

Figura 66

Grafica de efecto parcial de temp_low S(1)



Nota. Elaboración propia.

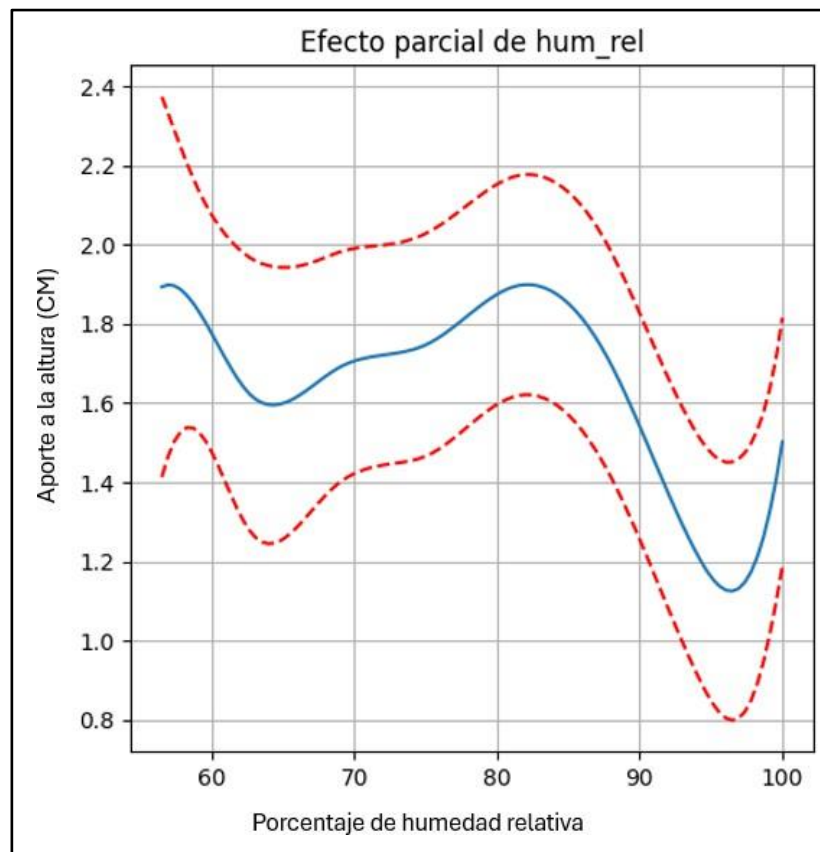
En la figura 66 se muestra la gráfica de la temperatura mínima que se presenta durante la noche que es las horas que presenta frío, la cual muestra picos y valles en donde el efecto en la planta puede ser positivo, pero en otros puede ser perjudicial afectando negativamente al crecimiento de la planta. Permite deducir que tener un rango moderado de temperatura mínima en momentos donde haga bastante frío.

El modelo al tener un EDoF alrededor de 8.5 permite una flexibilidad para poder capturar la complejidad de este efecto sin sobre ajustarse al ruido, ya que la significancia

estadísticamente alta, lo que sugiere que la variación de esta variable puede tener un impacto importante en el crecimiento, aunque para sugerir un rango se toma el que mayor aporta dando de temperatura mínima de 2.5 a aproximadamente 6 grados Celsius con mayor aporte a la altura.

Figura 67

Grafica de efecto parcial de hum_rel S(2)



Nota. Autoría propia.

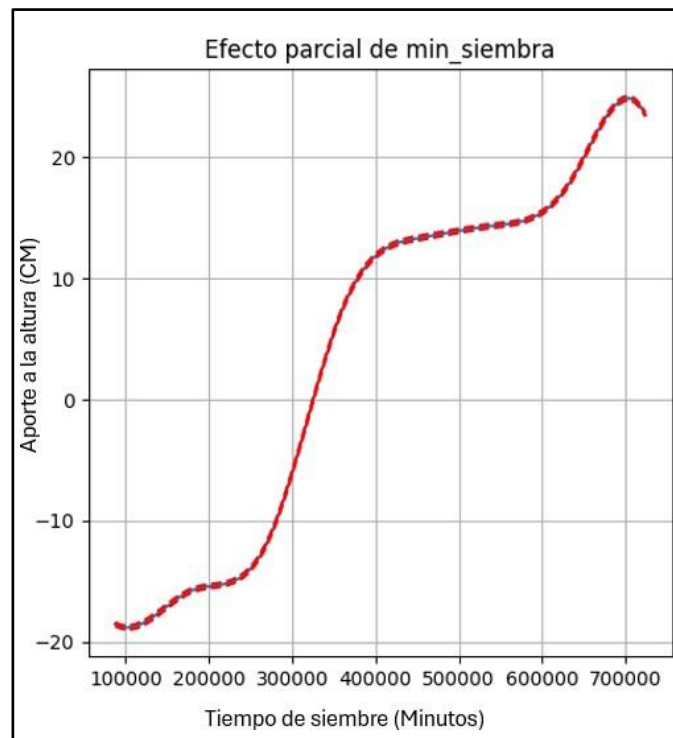
La figura 67 muestra la gráfica de la variable de humedad relativa en el aire, deduciendo por su suave variación con un pico que representa que la humedad capturada es un nivel adecuado en donde la planta crece de manera adecuada.

Una humedad del ambiente que este entre los rangos de 75% a 90% según el modelo sería lo que le aportaría a un correcto crecimiento de la planta, de lo contrario ya el efecto no sería tan positivo y aportaría poco a la variable objetivo.

Por último, para poder tener una humedad adecuada se pueden definir rangos por medio de estrategias de riego o una circulación de ventilación por medio de actuadores que ayuden a estabilizar la humedad el ambiente, para poder tener un crecimiento adecuado.

Figura 68

Gráfica de efecto parcial de $min_siembra S(3)$



Nota. Elaboración propia.

En la figura 68 se muestra la gráfica de los minutos transcurridos desde la siembra a de las plantas, la curva muestra un crecimiento claro y sostenidos, con incrementos más pronunciados conforme pasa el tiempo. Una de las principales razones de esto es que el

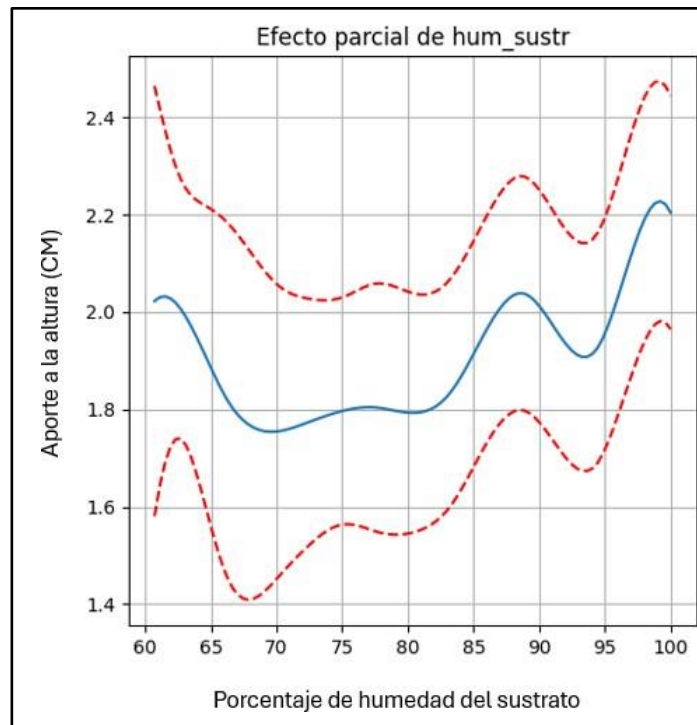
crecimiento de cualquier planta se correlaciona positivamente con el transcurrir del tiempo, ya que se acumula biomasa.

Basándonos en el **EDoF** alto permite tener presente la claridad del modelo para reconocer el paso del tiempo como un motor principal del crecimiento de la planta, además de que esta variable abarca un rango amplio lo que ayuda que el modelo capte el incremento del crecimiento de la planta.

Una medición precisa de esta variable permite tener una mejor predicción para la altura de la planta, abriendo paso a el uso de estrategias de prevención de enfermedades o intervenciones en fases críticas del crecimiento basadas en tiempo transcurrido después de la siembra.

Figura 69

Gráfica de efecto parcial de hum_sustr S(4)

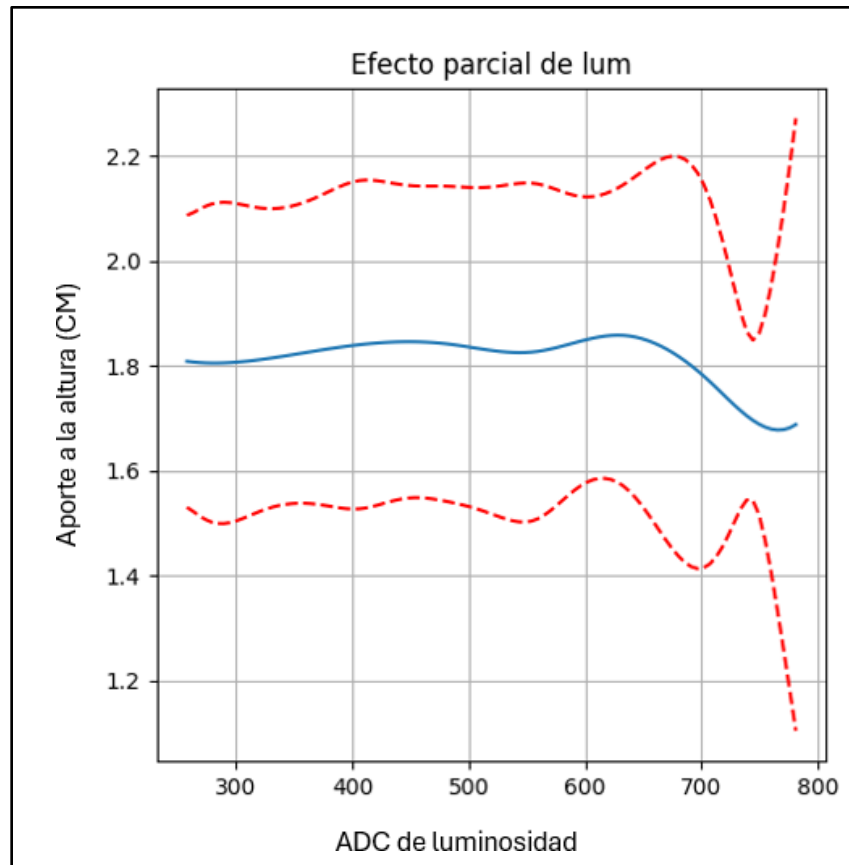


Nota. Autoría propia.

En la figura 69 se muestra la gráfica de la variable de humedad de sustrato, la cual muestra oscilaciones con efectos positivos como negativos, lo que sugiere que deben de existir rangos ideales para esta variable, determinando la no linealidad de esta variable, rangos adecuados como se muestra en la gráfica de valores mayores a 85%.

Por último, un **EDoF** intermedio (7.8) indica que el modelo ajusta una curva flexible, pero con una variabilidad que refleja la complejidad de la respuesta de la planta ante el cambio en la humedad del sustrato.

Se destaca la importancia de mantener una medida de humedad de sustrato adecuada por medio de riegos programados sin ocasionar saturación del sustrato apoyado en sistema de drenaje para poder contribuir a un equilibrio e ir en la dirección de un crecimiento adecuado de la planta.

Figura 70*Gráfica de efecto parcial de lum S(5)**Nota. Autoría propia.*

En la figura 70 se muestra la gráfica de la variable de luminosidad la cual muestra una tendencia casi que plana lo que indica que hay poca variación en el efecto parcial, esto se debe a que la luminosidad no varía de manera drástica en el dataset.

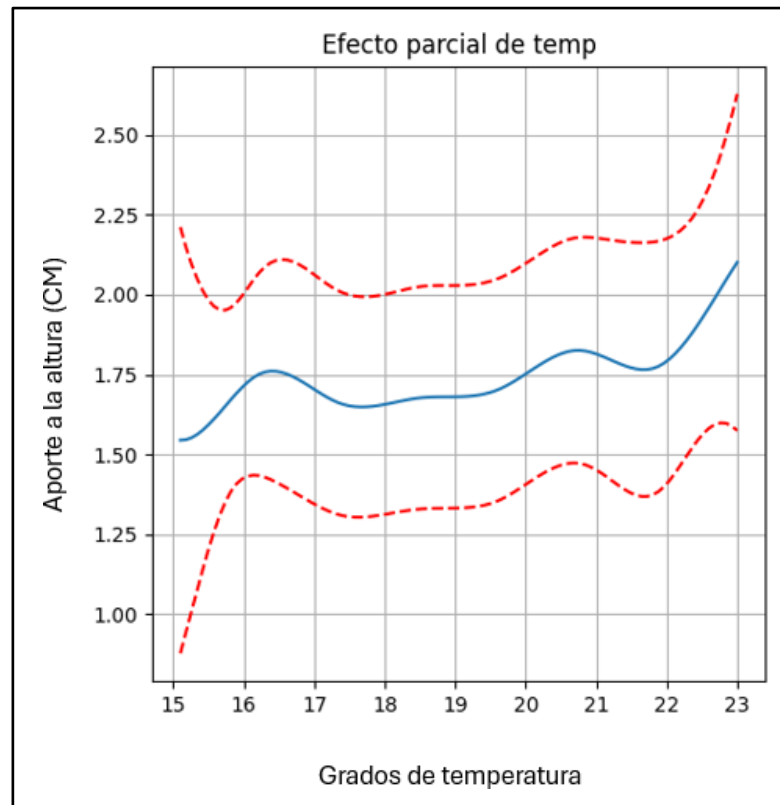
Con el **p-value** que arroja el modelo determina según el modelo que esta variable no tiene un mayor impacto en el crecimiento de la planta, al menos en este contexto haciendo que su efecto aislado sea difícil de detectar.

Es posible que la combinación de otras variables (como la temperatura y la humedad) interactúe con la luz de forma que su efecto se vea modulada en el ajuste global

del modelo. Estas medidas se deben a el rango de valores del sensor encargado de esta medida, aunque no se descarta que la luminosidad juega un papel fundamental para la fotosíntesis, al trabajar con estos datos el modelo la toma como no relevante en el crecimiento de la planta.

Figura 71

Grafica de efecto parcial de temp S(6)



Nota. Autoría propia.

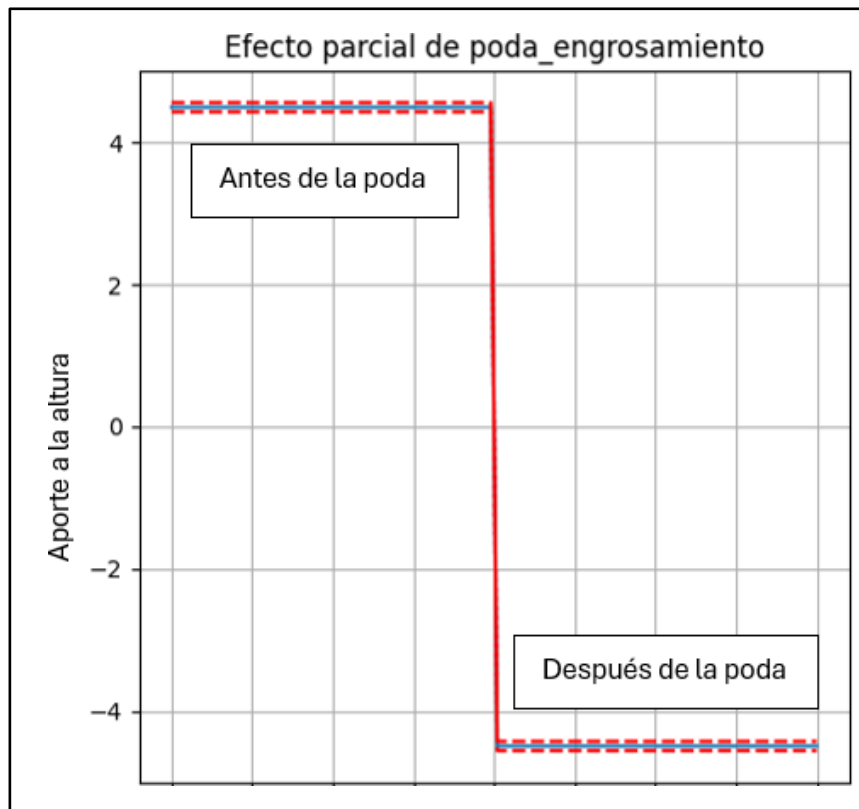
En la figura 71 se muestra el comportamiento del promedio entre la temperatura máxima y mínima, lo que se presenta de manera ondulada con uno o dos picos y valles, muy similar a lo que se observó en la figura X de temp_high y en la Figura X de temp_low, pero representado en un comportamiento promedio, determinando que el promedio de 16 a 22 grados Celsius.

Con los resultados obtenidos en el modelo de **EDoF** de 7.0 indica que se captura sin sobre ajustarse, reflejando que la temperatura promedio tiene un efecto moderadamente flexible pero predecible.

La temperatura al tener un factor de importancia alto es crucial manejar estrategias que permitan tener los ambientes con rangos adecuados, por medio de calefacción, enfriamiento o ventilación permitiendo un correcto crecimiento de la planta, en general la integración de la temperatura promedio con otros índices de los ambientes podría mejorar la calidad de la predicción.

Figura 72

Gráfica poda_engrosamiento (Variable categórica poda.f(7))



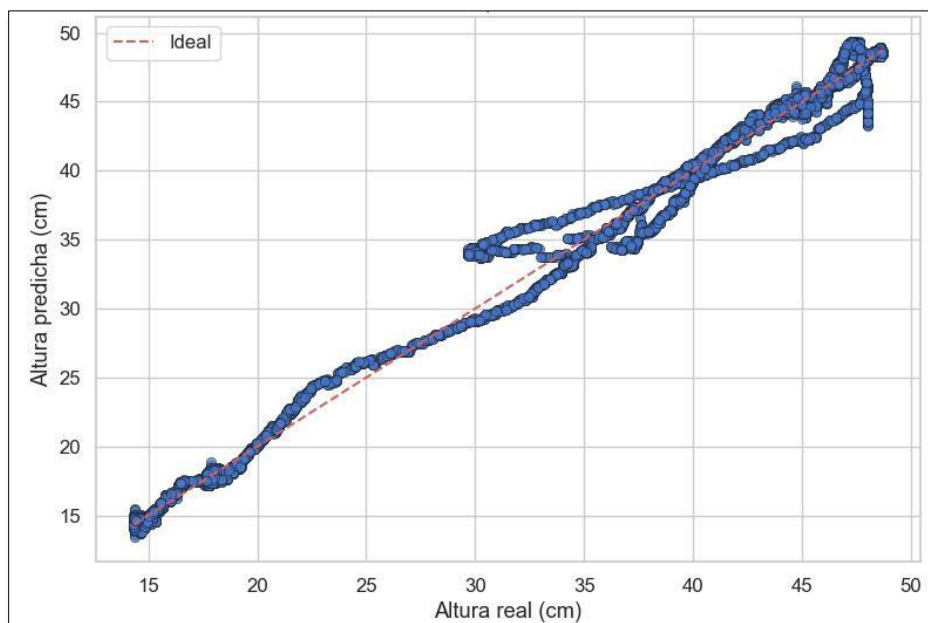
Nota. Elaboración propia.

En la figura 72 se muestra la gráfica de la variable poda de engrosamiento, al tratarse de una variable categórica se establece una categoría de referencia el efecto 0 sin poda y el efecto 1 con poda, haciendo un desplazamiento hacia arriba o hacia abajo. La práctica de realizar la poda de engrosamiento permite cambiar la distribución de recursos de la planta desarrollando basales más gruesos y estables.

La diferencia que arroja el modelo de coeficiente final de 18.2338 como suma de términos evidencia que al aplicar la poda se modifica la altura final de la planta, ya que el crecimiento se estimula de manera vertical o se redirecciona la energía de la planta, dependiendo de esto se puede evaluar el impacto que esta pueda tener en el crecimiento de la planta.

Figura 73

Gráfica de altura predicha vs altura real

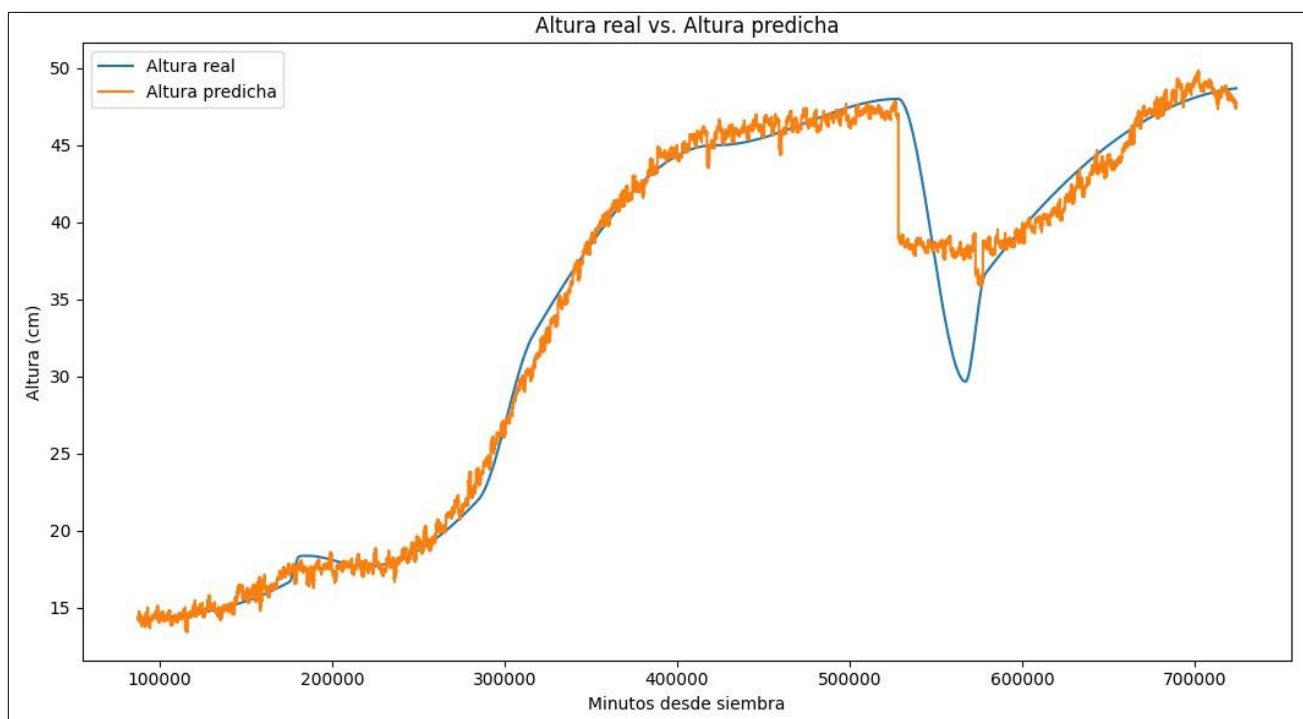


Nota. Elaboración propia.

En la Figura 73 se muestra una línea roja de tendencia creciente y lineal, la cual se toma como referencia para representar el comportamiento general de la altura de la planta. Además, se visualizan los puntos de predicción generados por el modelo GAM (Modelo Aditivo Generalizado). Este modelo demuestra una alta capacidad para ajustarse de manera flexible a la forma de la curva observada, lo que permite capturar mejor las variaciones reales del crecimiento sin caer en el sobreajuste.

Figura 74

Gráfica de tiempo vs altura predicha y altura real



Nota. Autoría propia.

En la gráfica de la figura 74 se observa una coherencia entre la altura real de la planta y las predicciones generadas por el modelo GAM a lo largo del tiempo transcurrido desde la siembra este medido en minutos, el modelo captura de manera adecuada el crecimiento acelerado, estabilización e incluso acercándose al descenso de la altura real,

está relacionada con la variable categórica de la poda. Las diferencias ocasionales entre las líneas pueden deberse a una variabilidad natural.

En conjunto el modelo presenta una predicción aceptable basado en dinámicas no lineales del crecimiento vegetal de la planta referente a la variable objetivo que es la altura de la planta.

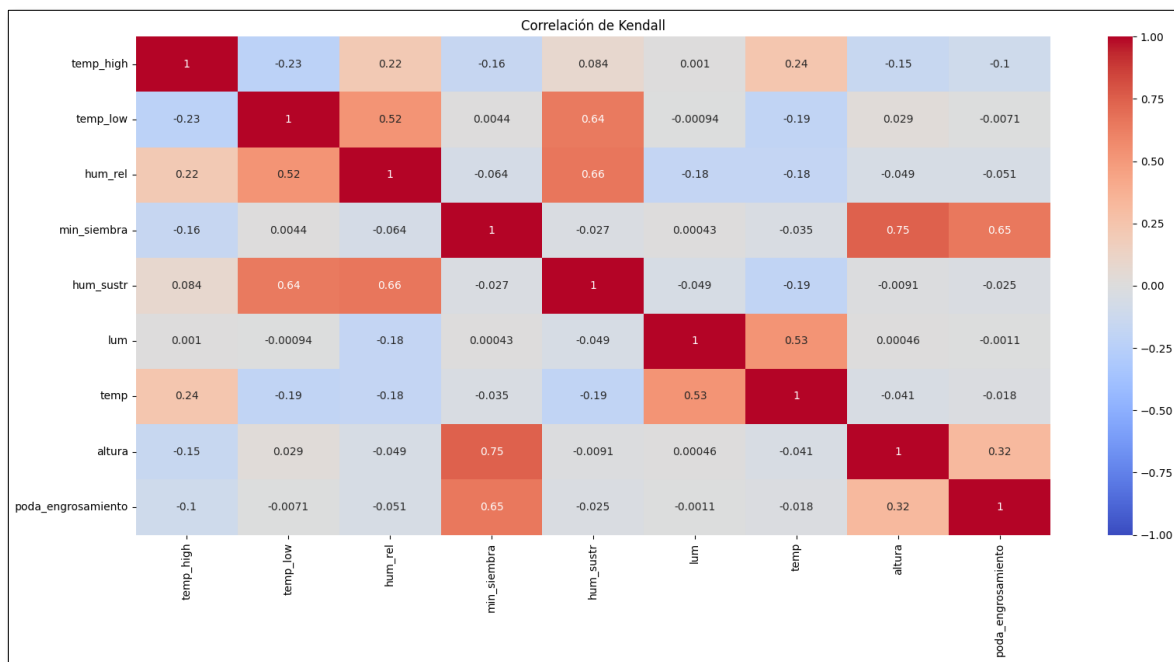
2.9.3. Correlación de Kendall para Determinar la Relación entre las Variables

Evaluada por GAM

La correlación de Kendall, también conocida como tau de Kendall, es una medida no paramétrica que evalúa la asociación entre dos variables ordinales, considerando la concordancia y discordancia entre pares de observaciones. A diferencia de otras correlaciones, como la de Pearson, no requiere que los datos sigan una distribución normal, lo que la hace adecuada para datos no lineales o con empates (Kendall, 1938; Sheskin, 2020).

Adicionalmente se hizo uso de la correlación de Kendall para determinar la relación de las variables, se decidió usar esta correlación dado que no todas las variables contaban con una distribución normal.

Figura 75
Matriz de correlación Kendall



Nota. Elaboración propia.

En la figura 75 se muestra la matriz de mapa de calor obtenida desde el programa la cual muestra que no hay una relación significativa entre las variables analizadas, ya que la mayoría de los coeficientes se encuentran con valores bajos, tanto positivos como negativos, de lo cual se deduce que las variables no están fuertemente relacionadas entre ellas, indicando que no se agrega mucha redundancia de información al modelo.

La baja interdependencia entre las variables permite ver que cada una aporta información relevante y única en análisis ya que los coeficientes no son ni iguales ni mayor a 1 y presenta valores bajos, permitiendo complementar el análisis determinado por el modelo GAM en donde mejora su interpretabilidad, evitando que este caiga en sesgos o errores por redundancia entre los predictores.

2.9.4. Conclusiones de los Resultados Obtenidos del Modelo GAM

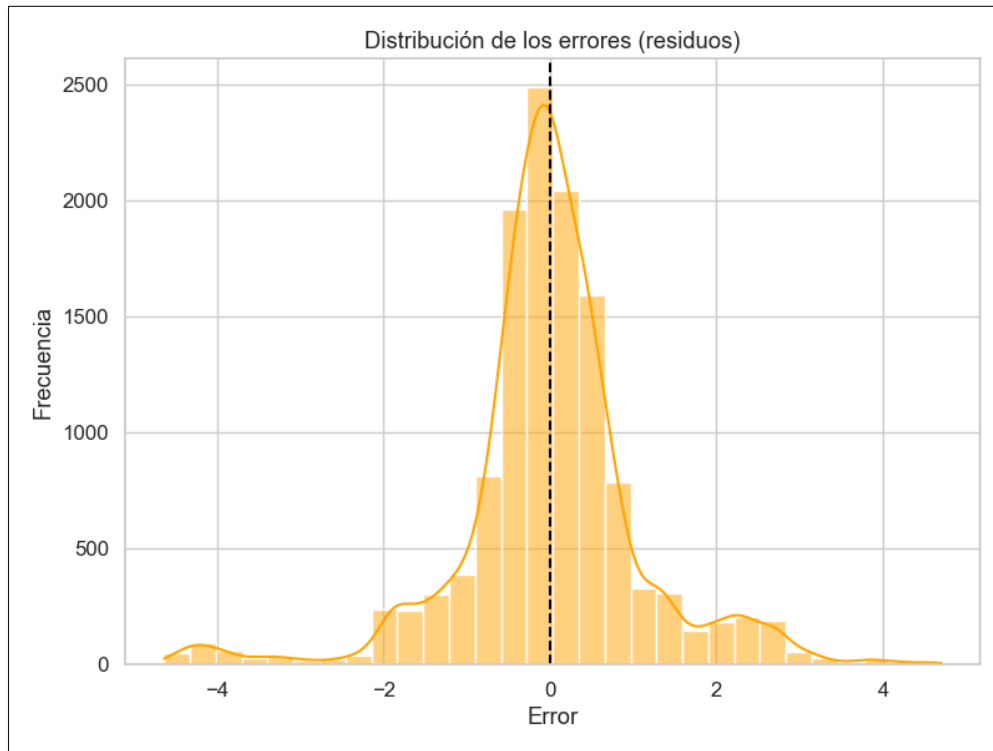
Finalmente, el análisis detallado muestra que el modelo GAM es capaz de capturar la complejidad del crecimiento de la planta mediante la integración de múltiples factores ambientales, temporales y agronómicos, cada variable muestra un patrón específico.

- **Variables continuas (temp_high, temp_low, hum_rel, hum_sustr, temp):** revelan relaciones no lineales con efectos en rangos específicos, muy importante para poder definir las condiciones ideales en un cultivo artificial, también permitiendo inferir que los extremos en cada una de estas pueden traer efectos negativos.
- **Tiempo (Min_siembra):** confirma que el crecimiento está directamente asociado al avance del tiempo del ciclo de vida de la planta.
- **Variable categórica (Poda_engrosamiento):** Resalta el impacto de las prácticas de manejo que se tengan en el cultivo, siendo estas potencialmente ajustables para optimizar la altura de la planta o cambiar la dirección de esta.

Por ultimo las métricas obtenidas en el modelo tales como el El alto coeficiente de determinación (R^2) y el bajo error (RMSE) indican que, el modelo captura de manera solida la variabilidad de la altura de la planta, esto no solamente permite entender cómo influye cada variable de manera “aislada ” y entender mejor los rangos en los que las condiciones son favorables o desfavorables para el crecimiento, sino que también poder diseñar estrategias para poder intervenir y llegar a optimizar el crecimiento de la planta basando en variables ideales y prácticas de correcto manejo.

Figura 76

Gráfica de distribución de los errores (Residuos)



Nota. Autoría propia.

La gráfica de la figura 76 muestra que la mayoría de los residuos se concentran alrededor de cero, lo cual indica que el modelo predice la altura de manera bastante precisa y sin sesgo sistemático; es decir, no tiende a sobrestimar ni subestimar de forma consistente. La forma simétrica y casi en forma de campana de la distribución (con colas que se extienden de forma moderada hacia los extremos) sugiere que los errores se comportan de manera similar a una distribución normal, apoyando el supuesto de homocedasticidad y normalidad en los residuos que se suele asumir en estos modelos. Las colas evidencian la presencia de algunos outliers o puntos atípicos, pero en general, estos casos son pocos y no afectan la robustez del modelo.

2.9.5. Modelos de Curva Sigmoide (S)

Los modelos de curva S o de función sigmoide son aquellos que su curva busca a ser igual a una S por lo tanto en el transcurso del estado del arte se pudo evidenciar que varios estudios presentaban el crecimiento vegetal, animal y bacteriano de esta manera, por ende, se buscó tener también la práctica con los datos obtenidos de la planta por medio de dos funciones la Gompertz y la Logística que se explican a continuación

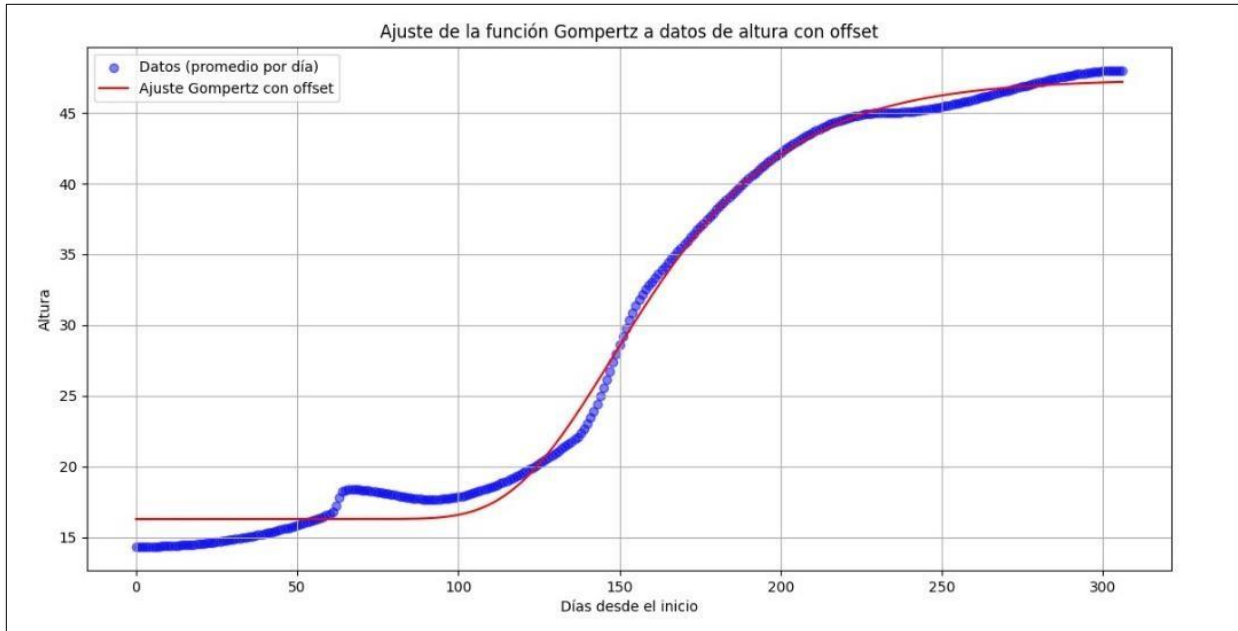
2.9.6. Función Gompertz

Se busco describir el crecimiento de la planta conforme el paso del tiempo con algún modelo logístico de los que se pudieron obtener en el transcurso de la investigación optando por el de función Gompertz el cual según Winsor “es un modelo matemático utilizado para describir procesos de crecimiento biológico y poblacional que presentan una desaceleración progresiva con el tiempo” (Winsor, 1932).

Para el ajuste de esta función se realizó con datos del 1 de enero de 2024 hasta el 2 de noviembre del 2024, justo antes de realizar la poda de engrosamiento de la planta, en la cual la curva de crecimiento se muestra cercana a una S, por ende se realizó una función de ajuste a los datos de altura, implementado un offset para que el modelo empiece con valor diferente a 0 cm ya que los datos están desde la fecha de 1 de enero del 2024 donde las plantas ya tenía una altura diferente a 0 cm, esto permite que el modelo con un parámetro adicional pueda ajustarse de la manera más adecuada posible a los datos reales como se muestra en la figura 77. Estos modelos al estar limitados a la forma S no podrían captar el evento de la poda, por esta razón el ajuste se limita al 2 de noviembre.

Figura 77

Ajuste de función Gompertz a datos de altura con offset



Nota. Autoría propia.

Para la estimación de los parámetros iniciales se tienen de la gráfica también se emplean los mismos datos, teniendo presente que se emplean las siguientes librerías en el desarrollo del código se emplearon las siguientes librerías:

- **Numpy**: para la manipulación numérica.
- **Matplotlib.Pyplot**: para graficas.
- **Scipy.optimize.curve_fit**: para ajustar funciones no lineales a los datos.
- **Pandas**: para realizar el cargue y manipulación de archivos csv.

Determinado que x_data es el índice que representa los días (eje x) y y_data los valores promedio de altura por día (eje y), Para la codificación de esta función se tiene la siguiente función que recibe los parámetros como se describe.

```
def gompertz(x, L, k, x0, y0):
    return y0 + L * np.exp(-np.exp(-k * (x - x0)))
```

Donde L es el incremento de la altura, K es la tasa de crecimiento, X0 es el punto de inflexión (donde el crecimiento cambia de rápido a lento) y Y0 el offset o el valor base de inicio para que el modelo no empiece desde 0. La estimación de los parámetros iniciales se realiza con lo siguiente.

```
y_max = y_data.max()
y_min = y_data.min()
p0 = [y_max - y_min, 0.1, x_data.mean(), y_min]
```

- **L:** diferencia entre valor máximo y mínimo.
- **k:** valor bajo (crecimiento suave).
- **x0:** valor medio del eje X.
- **y0:** valor mínimo (altitud base).

Y se ajusta la curva de Gompertz, con ayuda de `curve_fit` de *Sklearn*, teniendo presente la siguiente forma

```
popt, _ = curve_fit(gompertz, x_data, y_data, p0=p0, maxfev=10000)
L, k, x0, y0 = popt
```

Donde:

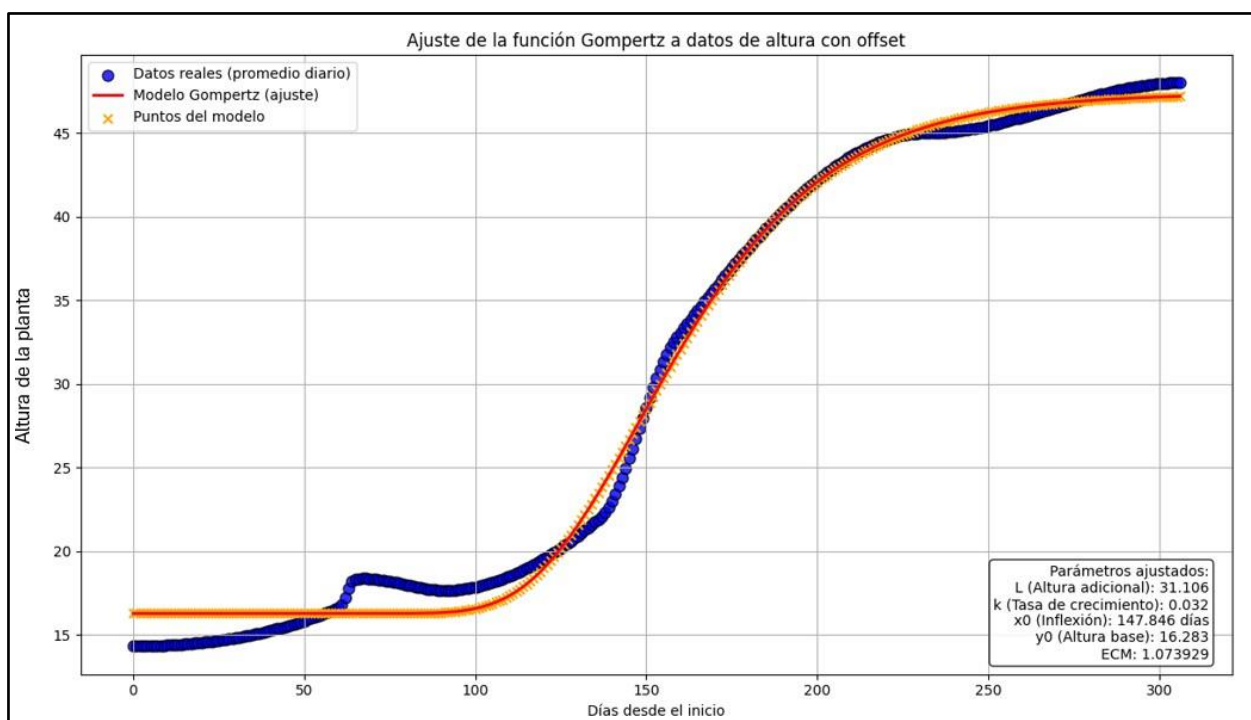
- **curve_fit:** busca los mejores valores de los parámetros para que la curva Gompertz se ajuste lo mejor posible a los datos.
- **maxfev=10000:** aumenta el número de intentos para encontrar una buena solución.

Como métricas se hace predicción y error, por medio de `y_pred` para los valores predichos y MSE para el error del modelo, dando como resultado la siguiente grafica

completa con los datos enriquecidos y mostrando la interacción de cada uno de los datos de manera que sea más fácil de interpretar visualmente como se muestra en la figura 78.

Figura 78

Ajuste de función Gompertz a datos de altura con offset y datos enriquecidos



Nota. Elaboración propia.

En la figura 78 se muestra el ajuste que tiene el modelo **Gompertz** respecto a los datos reales, dando como resultado los siguientes parámetros

Tabla 32

Resultados de ajuste del modelo Gompertz

Parámetros ajustados (escala original)	
Altura adicional (L):	31.106
Tasa de crecimiento (k):	0.032
Parámetro de inflexión (x0):	147.846 (días)

Valor base (y0):	16.283
Error cuadrático medio:	1.073929

Nota. Autoría propia.

Basado en estos datos se puede construir la siguiente función

$$y(x) = 16.283 + 31.106 * e(-e(-0.032(x - 147.846)))$$

Basado en todo lo visto anteriormente se puede describir que el crecimiento de la planta tiende a tener un comportamiento en **curva S (sigmoide)** con fase de aceleración seguida de una estabilización hacia una altura máxima.

Donde el inicio del crecimiento es lento durante aproximadamente los primeros **50 a 100 días**, luego tiene un crecimiento acelerado cerca a los **100 días** y en aproximadamente a los **190 días** se vuelve exponencial, alcanza su máxima velocidad en el día **147.8** el cual se interpreta con el **punto de inflexión (x0)**, por último, después del **día 190** el crecimiento se desacelera progresivamente hasta buscar estabilizarse en la altura máxima que se determinaría **16.283+31.106= 47.386 cm**

2.9.7. Función Logística

La función logística que se presentó en el estado del arte al igual que la de Gompertz permite modelar curvas de crecimiento no lineales que tienden a tener una forma S.

Para el ajuste del modelo se siguen los mismos pasos del Gompertz a excepción de que se define la función logística con offset de la siguiente manera

```
def logistica(x, L, k, x0, y0):
    return y0 + L / (1 + np.exp(-k * (x - x0)))
```

- **L**: altura máxima adicional que se alcanza (amplitud del crecimiento).
- **k**: tasa de crecimiento.
- **x0**: punto medio de crecimiento (inflexión).
- **y0**: valor base o altura inicial.

```
y_max = y_data.max()
```

```
y_min = y_data.min()
```

```
p0 = [y_max - y_min, 0.1, x_data.mean(), y_min]
```

- **L**: diferencia entre altura máxima y mínima.
- **k**: tasa inicial asumida.
- **x0**: punto medio del eje x.
- **y0**: altura más baja registrada.

```
popt, _ = curve_fit(logistica, x_data, y_data, p0=p0, maxfev=10000)
```

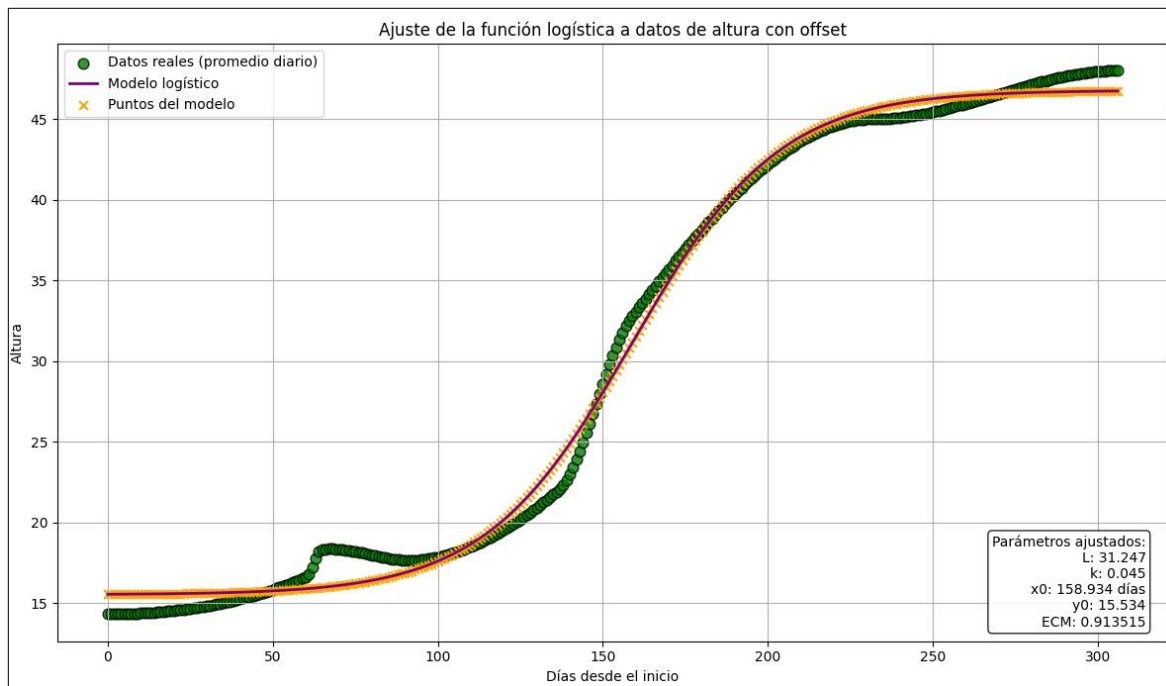
```
L, k, x0, y0 = popt
```

- **curve_fit**: encuentra los mejores parámetros para que la curva logística se ajuste a los datos.
- **maxfev=10000**: número máximo de evaluaciones para asegurar convergencia.

Se desempaquetan los parámetros ajustados.

Figura 79

Ajuste de función logística a datos de altura con offset y datos enriquecidos



Nota. Autoría propia.

En la figura 79 se muestra el resultado de cómo se ajusta la función logística a los datos reales, además de los diferentes valores que arroja la función para ajustarse y tener la función final.

Tabla 33

Resultados de ajuste del modelo logístico

Parámetros ajustados:	
Altura adicional (L):	31.247
Tasa de crecimiento (k):	0.045
Parámetro de inflexión (x0):	158.934 días
Valor base (y0):	16.283

Error cuadrático medio:	0.913515
Valor base (y0):	15.534

Nota. Elaboración propia.

En la tabla 33 se muestran los resultados finales arrojados para poder construir la función logística del modelo

Basado en estos datos se puede construir la siguiente función

$$y(x) = 15.534 + \frac{31.247}{(1 + e(-0.045 * (x - 158.934)))}$$

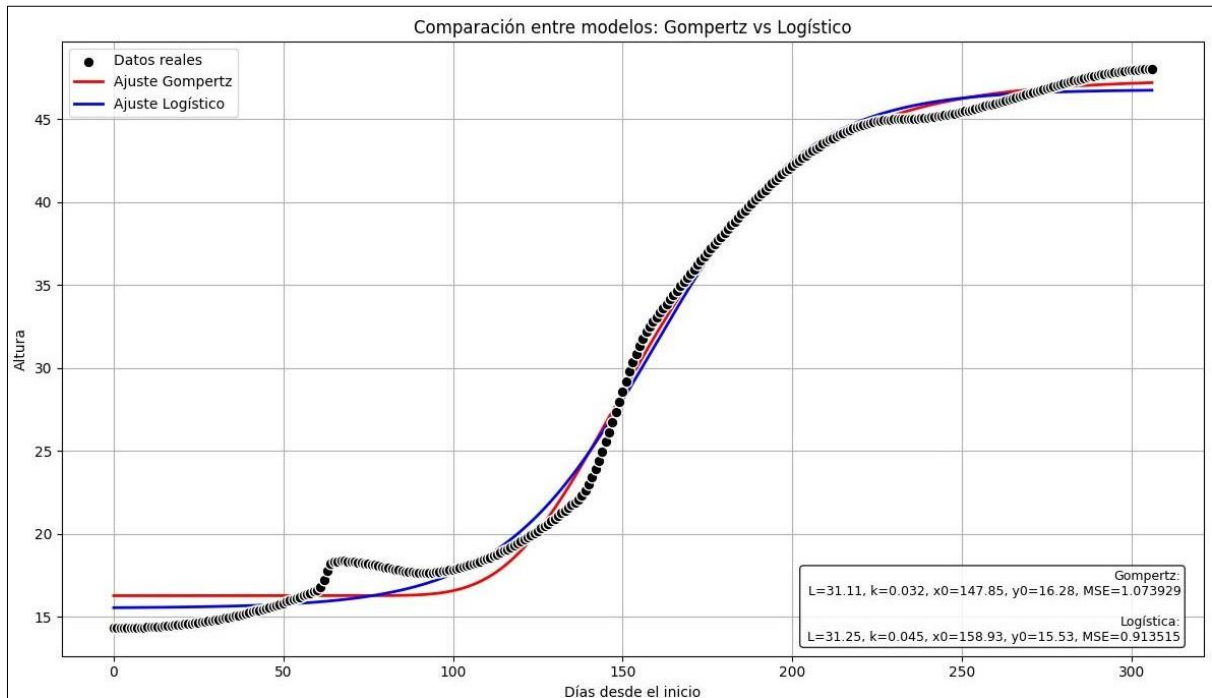
La función logística describe el crecimiento desde el inicio el incremento es lento lo cual es la fase de establecimiento, la fase inicial es alrededor de los 0 días hasta aproximadamente **80 – 100 días** ya que la curva crece, pero lo hace con lentitud, luego pasa a la fase de crecimiento acelerado que esta entre los **120 y 200 días** ya que crece notablemente es decir que el crecimiento es más rápido, la fase de saturación es a partir de aproximadamente a los **200 días** acercándose a un valor máximo de altura.

El MSE dado por esta grafica es aproximadamente de **0.9135**, el cual indica que los valores predichos se acercan a la real y se ajusta bastante cerca

2.9.8. Comparativa entre Gompertz y Logística

Figura 80

Gráfica de comparativa entre Gompertz y Logística



Nota. Autoría propia.

En la figura 80 se muestra la comparativa de las dos curvas, de color azul la logística, de color naranja la Gompertz y de color negro los valores reales, en donde basado en el MSE la que mejor se ajusta es la logística brindando un mejor análisis del crecimiento de la planta, además de poder tener una cercanía a los datos reales, en donde el ascenso paulatino que se vuelve más pronunciado hasta $x=158.9$ días aproximadamente, también que el crecimiento puede acelerarse a la asíntota de aproximadamente 46.8 cm indicando una posible altura máxima de la planta.

2.10. Resultados

En este apartado se presentan los resultados obtenidos durante el desarrollo del proyecto, los cuales se estructuran en función de los objetivos planteados inicialmente al momento de radicar la propuesta. El propósito es evidenciar tanto el avance en el proceso investigativo como los productos finales entregables, destacando particularmente el desarrollo de un módulo de reportes.

El objetivo general consistía en desarrollar un modelo matemático que permitiera describir el crecimiento de la planta de arándano Biloxi, mediante la comparación de variables ambientales. Para alcanzar este propósito, se definieron una serie de objetivos específicos, cada uno de los cuales guio las etapas del proyecto y cuya ejecución se refleja en los resultados obtenidos.

En primer lugar, se llevó a cabo la integración y análisis de los datos recolectados sobre el entorno y el crecimiento de la planta de arándano Biloxi. Esta fase fue fundamental para comprender las condiciones ambientales que inciden en el comportamiento de la planta y para disponer de una base de datos sólida que alimentara el modelo matemático, esto con el fin de poder construir un dataset que sirviera para poder alimentar a los modelos.

Resultados: Integración de datos

Para el resultado de los datos obtenidos se estudiaron soluciones para poder generar datos sintéticos basados en masas pequeñas de datos reales, ya que se presentaron algunas complicaciones para obtener un volumen de datos adecuado para alimentar los modelos. Al llevar a cabo la investigación se tuvieron presentes como plan A los modelos **ARIMA** y

SARIMA, y como plan B una red neuronal, con el paso del desarrollo de esto se optó por la red neuronal cuya arquitectura se explica en el apartado de datos, con esta red se obtuvieron datos sintéticos que permitieron adaptarse al comportamiento de estos datos, generando un dataset amplio que permitió tener un volumen de datos adecuado para obtener resultados del modelo y como cada una de las variables estudiadas como se muestra en la siguiente figura.

Figura 81

Información sobre el dataset obtenido

```
Archivo cargado: C:/Users/jhona/Desktop/Modelos/dataset_con_poda.csv  
  
Cantidad de datos (filas): 63668  
Cantidad de variables (columnas): 10  
  
Variables:  
timestamp  
temp_high  
temp_low  
hum_rel  
min_siembra  
hum_sustr  
lum  
temp  
altura  
poda_engrosamiento
```

Nota. Autoría propia.

Posteriormente, se procedió al desarrollo de los modelos matemáticos propiamente dichos, utilizando herramientas informáticas como Python y sus librerías que se describen en el capítulo de modelos. Estos modelos fueron diseñados para describir el crecimiento de la planta a partir de los datos recolectados, permitiendo identificar el aporte que tenía cada una de las variables en la altura de la planta.

Resultados: Desarrollo modelo matemático

Para el desarrollo de tuvieron presente 3 modelos, dos de curva sigmoide y uno GAM, esto con el fin de poder tener un análisis del comportamiento de las variables estudiadas respecto a la variable objetivo que en este caso es la altura.

Por un lado, el modelo GAM permitió obtener rangos en donde cada una de las variables estudiadas tuvieran un buen aporte al crecimiento de la planta, permitiendo diseñar estrategias en donde por medio de controles se puedan llegar a estos rangos y así poder tener una planta en un correcto crecimiento, los rangos donde cada variable aporta más positivamente a la altura fueron los siguientes, aunque los efectos no sean monotónicos:

- 18 a 23 grados Celsius para la temperatura alta durante el día.
- 2.6 a 6 grados Celsius para la temperatura mínima, aplicada a horarios nocturnos.
- 75 a 90 % de humedad relativa en el aire.
- Mayor a 85 % en la humedad del sustrato de la planta.
- 18 a 23 grados Celsius la temperatura del ambiente del sistema en donde se encuentre la planta.

Esto permite tener un análisis más detallado del comportamiento individual de las variables estudiadas en el ambiente que se tenga.

El modelo GAM sugiere que la temperatura máxima diaria más favorable para el arándano Biloxi oscila aproximadamente entre 18 °C y 23 °C. Dentro de este rango, el modelo muestra un efecto positivo máximo sobre la altura, disminuyendo fuera de dichos

límites. Agronómicamente, este resultado es coherente con las condiciones templadas que favorecen el crecimiento del arándano. Biloxi es un cultivar adaptado a climas moderados, más tolerante al calor que los arándanos de mayor requerimiento de frío, pero aun así prospera mejor con temperaturas diurnas relativamente suaves (Ministerio de Agricultura, Ganadería y Alimentación de Guatemala, 2024), (Hydro Environment. s.f.). En la Sabana de Bogotá (Colombia), por ejemplo, las temperaturas máximas típicas rondan 18–20 °C, condición en la cual se ha logrado cultivar exitosamente Biloxi (Casas Rodríguez, 2017). Esto concuerda con el rango que identifica el modelo.

Es importante aclarar que la temperatura máxima analizada proviene de registros meteorológicos a nivel de ciudad y no de sensores localizados en el cultivo, cómo se explica en el capítulo de integración de datos. Incluir este dato externo en el modelo GAM pudo ser discutible, pero se consideró que, al ser un dato usado por los generadores de datos sintéticos de temperatura esta variable también representa en parte el microclima del experimento. No obstante, el hecho de que el rango óptimo identificado (18–23 °C) coincide con lo esperado en un clima altoandino moderado sugiere que el modelo, aun con este dato, capturó una señal consistente con la literatura.

Según el modelo GAM, la temperatura mínima diaria óptima se sitúa aproximadamente entre 2.5 °C y 6 °C. Es decir, noches frías, cercanas a los 5 °C, parecen asociarse a mejores resultados en el cultivo. En literatura se destaca que los arándanos necesitan acumular horas de frío (exposición a temperaturas bajas) para inducir una adecuada floración. En variedades southern highbush como Biloxi, este requerimiento es bajo pero existente: se estiman alrededor de 150 a 400 horas por debajo de ~7 °C para romper la dormancia de las yemas (Hydro Environment. s.f.). Esto explica por qué en

climas sin invierno marcado (tropicales de altitud) Biloxi aún responde positivamente a las noches frías recurrentes. De hecho, Biloxi ha sido reportado como “de bajo requerimiento de frío”, ideal para regiones de inviernos suaves (Hydro Environment. s.f.), lo cual coincide con la idea de que rangos mínimos en torno a 2–6 °C pueden ser suficientes para satisfacer parte de su necesidad de frío sin llegar a heladas severas.

Al igual que con la máxima, la temperatura mínima diaria analizada proviene de la estación meteorológica local y no de sensores propios del cultivo. Esto plantea las mismas consideraciones, pero el uso de ambos tipos de temperatura podría haber captar efectos distintos (la mínima externa reflejando frío general, y la temperatura ambiente reflejando condiciones microclimáticas). Los resultados concuerdan con la literatura en mostrar que noches frías benefician al arándano (Hydro Environment. s.f.).

La temperatura ambiente se refiere a la temperatura del aire medida directamente en el cultivo, mediante sensores. A diferencia de las “temperaturas extremas” diarias (máx/mín), esta variable proporciona el microclima real que experimentan las plantas durante el día. El modelo GAM encontró un determinado rango efectivo en torno a valores medios diarios de ~18–23 °C. Este rango representaría un clima fresco a templado, consistente con la suma de días moderadamente cálidos y noches frías ya discutidos. Desde el punto de vista agronómico, (entre 12 y 18 °C) es muy adecuado para el arándano. La literatura indica que el cultivo crece mejor en climas moderados, con temperaturas ni muy frías todo el día ni demasiado altas. Por ejemplo, un informe agroclimático de Guatemala sugiere un rango de 5 a 18 °C como óptimo diario para el desarrollo del arándano (Ministerio de Agricultura, Ganadería y Alimentación de Guatemala, 2024). Así, el rango

encontrado por el modelo para temperatura ambiente concuerda con las condiciones recomendadas.

En cuanto a la humedad relativa del aire desde el punto de vista agronómico, la literatura sugiere que un rango de 40% a 60% de HR es ideal para blueberries en invernadero, ya que equilibra la transpiración y reduce riesgos patogénicos. HR constantemente altas (>90%) facilitan la proliferación de hongos (Rankel, K, 2024), y aunque esta consideración exacta no tenga respaldo por parte del modelo, este sí sugiere un rango adecuado entre 75% y 90%, de hecho, reflejando una caída en el aporte a la altura en valores superiores a 90%.

Respecto a la humedad del sustrato, la literatura sugiere que un rango de 40% a 60% de humedad de sustrato es ideal para blueberries (Intagri S.C., 2018). Aunque en el experimento no se encontraron niveles de humedad tan bajos el modelo sugiere mayor aporte positivo a la altura de la planta en rangos entre 85% a 100%, pero revisando al detalle también representa un efecto no monotónico, teniendo otro máximo local sobre 60% de humedad de sustrato, un valor más adecuado según la literatura.

Los modelos de curva S, permitieron a diferencia del GAM identificar un comportamiento de los datos en general, arrojando datos que también son relevantes tales como la tasa de crecimiento, la fórmula de ajuste, valor de ajuste y otros aspectos. Primeramente, se usó un modelo gompertz el cual presento un ajuste a los datos aceptable arrojando los siguientes datos.

- **Altura Adicional (L):** 31.106
- **Tasa de crecimiento (k):** 0.032

- **parámetro de inflexión (x0):** 147.846 días
- **Valor base (y0);** 16.283
- **Error cuadrático medio: 1.073929**

Esto permitió tener una comprensión del crecimiento de la planta, además de tener un buen ajuste a los datos esto reflejado por el error cuadrático presentado.

Para poder buscar un modelo que pudiera brindar un mejor resultado se optó también por realizar un modelo logístico el cual presento un mejor ajuste a los datos, al igual que el Gompertz arrojaron datos de tasa de crecimiento y otros que se muestran a continuación.

- **Altura Adicional (L):** 31.247
- **Tasa de crecimiento (k):** 0.045
- **Parametro de inflexión (x0):** 158.934 días
- **Valor base (y0);** 16.283
- **Error cuadrático medio: 0.913515**

Como se puede ver el modelo logístico presento datos similares a los del modelo gompertz, aunque el modelo logístico se ajustó más a los datos arrojando un error cuadrático menor, lo que indica una mejora de precisión a la hora de poder arrojar los datos evaluados en el estudio descriptivo del crecimiento, de la misma forma esto se puede profundizar más en el capítulo de modelos en donde se muestran los resultados obtenidos de manera más detallada, al igual que la gráfica comparativa entre ambos modelos.

Como parte de la estrategia de visualización de los resultados, se desarrolló un módulo de software que integra el modelo matemático en un servicio web, facilitando así

su consulta, interpretación y análisis. Este módulo permite a los usuarios interactuar de manera sencilla con los datos y observar las proyecciones de crecimiento bajo diferentes condiciones ambientales

Figura 82

Resultado Modulo web



Nota. Elaboración propia.

En el módulo web se muestra todo aquello que se compone el modelo matemático desde fórmulas que se usaron, librería hasta los resultados obtenidos tal como se ve en la figura, es una pequeña muestra de lo que ofrece el módulo de reportes al usuario.

Finalmente, se realizaron pruebas funcionales y no funcionales sobre el módulo de software con herramientas como JMETER y SONARQUBE, esto reflejado en el apartado

2.3, asegurando que cumpliera con los requerimientos de calidad, y usabilidad previstos. Estas pruebas permitieron tener presentes detalles importantes como la seguridad, mantenibilidad y flexibilidad del proyecto.

El propósito de esta investigación y proyecto fue dar respuesta a la pregunta problema planteada inicialmente: **¿Cómo se puede describir el crecimiento de los arándanos Biloxi en un sistema cerrado y abierto de recirculación de agua, teniendo en cuenta las variables ambientales de su entorno?**

Para describir el crecimiento de los arándanos Biloxi en sistemas cerrados y abiertos de recirculación de agua, considerando las variables ambientales del entorno, se implementaron dos enfoques complementarios de modelado. El primero utilizó un modelo estadístico no paramétrico, específicamente un Modelo Aditivo Generalizado (GAM), que permitió analizar cómo cada variable ambiental influye individualmente en el crecimiento de la planta. Este modelo es capaz de capturar relaciones no lineales entre las variables y el crecimiento, proporcionando representaciones visuales claras de estos efectos. Además, al ser implementado en un archivo pkl.

El segundo enfoque se basó en modelos con funciones sigmoides, que, aunque no modelan explícitamente cada variable ambiental, describen el crecimiento de la planta en función del tiempo mediante una ecuación formal. Estos modelos permiten calcular parámetros como la tasa de crecimiento, ofreciendo una visión integral del desarrollo de la planta a lo largo del tiempo.

La combinación de estos modelos proporciona una comprensión del crecimiento de los arándanos Biloxi en diferentes sistemas de recirculación de agua, considerando tanto el

impacto individual de las variables ambientales como la dinámica temporal del crecimiento. Esto permite describir cómo las condiciones ambientales y los sistemas de cultivo afectan el desarrollo de las plantas, respondiendo así a la pregunta de investigación planteada.

De esta manera, el proyecto logró cumplir con los objetivos planteados, culminando con la entrega de un módulo de reportes que sintetiza el trabajo realizado y proporciona una herramienta útil para el seguimiento y análisis del crecimiento de la planta de arándano Biloxi en distintos entornos.

2.11. Conclusiones y Recomendaciones

Conclusiones

Como conclusión, es importante resaltar la necesidad de contar con un mayor volumen de datos reales para lograr una descripción más detallada del crecimiento en altura de la planta en función de las variables ambientales incluidas en el modelo. Disponer de datos reales no solo mejora la capacidad predictiva del modelo, sino que también permite comprender mejor el comportamiento y la interacción de las variables ambientales involucradas.

La generación de datos sintéticos representó un desafío debido al limitado volumen de información disponible. Aunque inicialmente se consideraron modelos como ARIMA y SARIMA para capturar tendencias estacionarias, se determinó que los datos eran insuficientes para obtener resultados confiables. Esta limitación condujo a la implementación de redes neuronales como alternativa, permitiendo modelar adecuadamente los patrones de los datos mediante una arquitectura específica. Gracias a este enfoque, fue posible generar información útil para alimentar los modelos. Asimismo, se abordó de forma

adecuada el tratamiento de valores nulos, para tener una mejor calidad de los datos utilizados.

Por otro lado, el modelo GAM, al ser una herramienta estadística más robusta, nos permitió realizar un análisis detallado del efecto que cada variable tiene sobre la altura de la planta. Las gráficas generadas evidencian su capacidad para modelar relaciones no lineales complejas, lo que lo convierte en una excelente opción frente a modelos más rígidos. Además, al analizar la correlación de Kendall entre las variables predictoras, se observó que estas no están fuertemente relacionadas entre sí, lo que facilita una interpretación más clara del aporte individual de cada una sobre la variable de respuesta (altura). Finalmente, el modelo GAM nos permitió identificar los valores óptimos de cada variable que favorecen un mayor crecimiento de las plantas.

Tabla 34

Rangos óptimos de las Variables

Valores óptimos obtenidos por medio de GAM

Valores óptimos	Variable dataset	Variable real
18 a 23 grados Celsius	temp_high	Temperatura alta
2.6 a 6 grados Celsius	temp_low	Temperatura mínima
75% a 90%	hum_rel	Humedad relativa en aire
Mayor de 85%	hum_sustr	Humedad de sustrato
	lum	Luminosidad
18 a 23 grados Celsius	temp	Temperatura media

Nota. Autoría propia.

En la tabla 31 se presentan los rangos que más aportan positivamente a la altura obtenidos a partir del análisis con el modelo GAM identificando valores óptimos de las variables.

Por otro lado, los modelos logísticos como el de Gompertz y el Logístico presentan un buen ajuste a los datos reales, especialmente el modelo logístico. Estos modelos ofrecen información general valiosa, como la tasa de crecimiento y el punto de inflexión de la curva, aspectos fundamentales para comprender el comportamiento del crecimiento.

Ambos modelos son herramientas importantes para describir de manera integral el crecimiento de la planta. Mientras los modelos logísticos permiten entender cómo varía el crecimiento a lo largo del tiempo mostrando un patrón que comienza lentamente, se acelera en una etapa intermedia y finalmente se desacelera, el modelo GAM ofrece un análisis más detallado al permitir evaluar el impacto individual de cada variable respecto a la altura. Esto facilita identificar los rangos específicos en los que dichas variables tienen un mayor efecto.

Recomendaciones

Una de las recomendaciones más relevantes para el éxito de este proyecto es asegurar la disponibilidad de datos de calidad. Es fundamental contar con un volumen adecuado de datos que permita alimentar correctamente a los modelos matemáticos. Estos datos pueden ser recolectados directamente a través de sensores instalados en el cultivo, teniendo una precisión y representatividad. Sin embargo, en caso de que la cantidad de datos reales sea limitada, se sugiere complementar el conjunto de datos mediante la generación de datos sintéticos utilizando modelos robustos de series temporales, como ARIMA, SARIMAX o redes neuronales más robustas. Estos modelos no solo permiten simular datos adicionales, sino que también ayudan a construir modelos matemáticos más sólidos, en los cuales las variables ambientales pueden ser incorporadas como variables externas, mejorando así la capacidad predictiva de los sistemas.

Por otra parte, se recomienda fortalecer el módulo web actual, especialmente en su capa de backend. Una mejora significativa sería la implementación de un CRUD (Crear, Leer, Actualizar y Eliminar) de datasets, lo que permitiría a los usuarios cargar nuevos conjuntos de datos, eliminar registros obsoletos y gestionar de manera más detallada el historial de datos recogidos del cultivo. Este tipo de funcionalidad no solo optimizaría el manejo de la información, sino que también facilitaría el mantenimiento y la escalabilidad de la herramienta a largo plazo.

Por otra parte, también se recomienda experimentar con otros modelos matemáticos más simples que permitan tener una mejor comprensión del comportamiento biológico del crecimiento de la planta. Adicionalmente, se propone expandir la aplicación de esta investigación hacia otras especies de bayas silvestres, así como a diferentes tipos de ambientes. Realizar pruebas en diversas condiciones climáticas y tipos de suelos permitirá evaluar la adaptabilidad y robustez del modelo propuesto, proporcionando información valiosa sobre su aplicabilidad en contextos más variados. Esta expansión contribuiría a validar la generalización de los modelos y a encontrar oportunidades de mejora para su implementación en la agricultura de precisión.

Finalmente, se subraya la importancia de que los usuarios de la herramienta tengan un entendimiento básico o intermedio de programación. El conocimiento en programación no solo facilitará la implementación y el uso adecuado del sistema, sino que también permitirá realizar ajustes o mejoras cuando sea necesario, asegurando un control más eficiente, limpio y flexible del software. De esta forma, se garantiza una mejor mantenibilidad y se potencia el aprovechamiento de las funcionalidades ofrecidas por la herramienta.

3. Bibliografia

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO₂ enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO₂. *New Phytologist*, 165(2), 351-372. <https://doi.org/10.1111/j.1469-8137.2004.01224.x>

Al-Chalabi, M. (2015). Vertical farming: Skyscraper sustainability? *Sustainable Cities and Society*, 18, 74-77.

Allison, P. D. (2001). *Missing data* (Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136). Sage Publications.

Archontoulis, S. V., & Miguez, F. E. (2015). Nonlinear regression models and applications in agricultural research. *Agronomy Journal*, 107(2), 786-798. <https://doi.org/10.2134/agronj2012.0506>

Altieri, M. A. (2004). Linking ecologists and traditional farmers in the search for sustainable agriculture. *Frontiers in Ecology and the Environment*, 2(1), 35-42.

Altieri, M. A., & Nicholls, C. I. (2017). Agroecology: a brief account of its origins and currents of thought in Latin America. *Agroecology and Sustainable Food Systems*, 41(3-4), 231-237.

Arroni del Riego, S. (2024). Generador de datos sintéticos tabulares basado en modelos de difusión [Tesis de maestría, Universidad Politécnica de Madrid]. Archivo Digital UPM. <https://oa.upm.es/82958/>

Balafoutis, A., Beck, B., Fountas, S., Vangeyte, J., Wal, T., Soto, I., Gómez-Barbero, M., Barnes, A., & Eory, V. (2017). Precision agriculture technologies positively contributing to GHG emissions mitigation, farm productivity and economics. *Sustainability*, 9(8), 1339.

Benages-Pardo, L., Sagiüés-Tanco, R., López-Nicolás, G., & Llorente, S. (2019). Generación de datos sintéticos con objetos de cocina para entrenar redes neuronales de convolución. En *XL Jornadas de Automática: libro de actas, Ferrol, 4-6 de septiembre de 2019* (pp. 170-177). <https://ruc.udc.es/dspace/handle/2183/23687>

Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). Precision agriculture and sustainability. *Precision Agriculture*, 5(4), 359-387.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th ed.). John Wiley & Sons.

Carnicer, J., Barbeta, A., Sperlich, D., Coll, M., & Peñuelas, J. (2013). Contrasting trait syndromes in angiosperms and conifers are associated with different responses of tree growth to temperature on a large scale. *Frontiers in Plant Science*, 4, 409.

<https://doi.org/10.3389/fpls.2013.00409>

Carvalho, S. M. P., Abi-Tarabay, H., & Heuvelink, E. (2019). Modeling plant growth and development. In *Encyclopedia of Sustainability Science and Technology* (pp. 1-26). Springer.

Casas Rodríguez, V. M. (2017). *Establecimiento preliminar de las condiciones agroclimáticas, zonas de adaptación y cultivares potenciales para el desarrollo del cultivo del arándano (Vaccinium corimbosum) en Colombia* (Trabajo de monografía, Universidad de Cundinamarca). Recuperado de <https://repositorio.ucundinamarca.edu.co/server/api/core/bitstreams/aec55f39-c3ec-4838-883a-54f7dafa931/content>

Ceccon, E., González, E. J., & Martorell, C. (2019). Is direct seeding a biologically viable strategy for restoring forest ecosystems? Evidences from a meta-analysis. *Land Degradation & Development*, 30(6), 744-757. <https://doi.org/10.1002/ldr.3254>

Chatfield, C. (2003). *The analysis of time series: an introduction* (6th ed.). Chapman and Hall/CRC.

Congreso de Colombia. (1993). Ley 99 de 1993: Creación del Ministerio del Medio Ambiente y regulación de la política ambiental. <https://www.minambiente.gov.co/wp-content/uploads/2021/08/ley-99-1993.pdf>

Congreso de Colombia. (1994). Ley 165 de 1994: Protocolo de Cartagena sobre seguridad de la biotecnología. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=37807>

Congreso de Colombia. (2002). Ley 740 de 2002: Acceso a recursos genéticos. https://www.minsalud.gov.co/Normatividad_Nuevo/Ley%200740%20DE%202002.pdf

Congreso de Colombia. (2012). Ley 1530 de 2012: Régimen de regalías para ciencia, tecnología e innovación.

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=47474>

Congreso de Colombia. (2012). Ley 1518 de 2012: Manejo integrado de plagas.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=47040>

Congreso de Colombia. (2017). Ley 1876 de 2017: Sistema Nacional de Innovación Agropecuaria. <https://www.suin-juriscol.gov.co/viewDocument.asp?ruta=Leyes/30034416>

Congreso de Colombia. (2020). Ley 2046 de 2020: Promoción de la Innovación Agropecuaria.
<https://www.suinjuriscol.gov.co/viewDocument.asp?ruta=Leyes/30039692>

Despommier, D. (2010). *The vertical farm: feeding the world in the 21st century*. Thomas Dunne Books.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. (2008). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5), 609-628.
<https://doi.org/10.1111/j.2007.0906-7590.05171.x>

Encolombia. (2021). Agricultura tradicional [Fotografía]. Encolombia.
<https://encolombia.com/wp-content/uploads/2021/01/Agricultura-Tradicional-696x320.jpg>

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Esteban Lasso, A., Martínez Toledo, C., & Perosanz Amarillo, S. (2023). Diseño de un modelo para generar datos sintéticos en investigación médica. *Dianas*, 12(1), e202303fp01. <https://ebuah.uah.es/xmlui/handle/10017/58141>

Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., Sadia, S., Nasim, W., Adkins, S., Saud, S., Ihsan, M. Z., Alharby, H., Wu, C., Wang, D., & Huang, J. (2017). Crop production under drought and heat stress: Plant responses and management options. *Frontiers in Plant Science*, 8, 1147. <https://doi.org/10.3389/fpls.2017.01147>

FAO. (2021). Conservation Agriculture. Food and Agriculture Organization of the United Nations. <http://www.fao.org/conservation-agriculture/en/>

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., & Basra, S. M. A. (2009). Plant drought stress: Effects, mechanisms and management. *Agronomy for Sustainable Development*, 29(1), 185-212. <https://doi.org/10.1051/agro:2008021>

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., & Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, 478(7369), 337-342

Freepik. (n.d.). Concepto de producción agrícola alimentaria: Silos, máquinas tractoras y sistema de riego [Fotografía]. Freepik. https://img.freepik.com/foto-gratis/concepto-produccion-agricola-alimentaria-silos-maquinas-tractoras-sistema-riego_342744-565.jpg

Flexas, J., Diaz-Espejo, A., Gago, J., Gallé, A., Galmés, J., Gulías, J., & Medrano, H. (2018). Photosynthetic limitations in Mediterranean plants: A review. *Environmental and Experimental Botany*, 153, 49-56.

G A N Pongdatu and Y H Putra 2018 IOP Conf. Ser.: Mater. Sci. Eng. 407 012153.

Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. *Science*, 327(5967), 828-831.

Gerhards, R., & Oebel, H. (2006). Practical experiences with a system for site-specific weed control in arable crops using real-time image analysis and GPS-controlled patch spraying. *Weed Research*, 46(3), 185-193.

Gómez-Hernández, L., & Martínez-Soto, G. (2020). Aplicación del análisis descriptivo en la evaluación del crecimiento vegetal bajo diferentes condiciones ambientales. *Revista de Botánica Experimental*, 45(2), 78-95.

Gratani, L., Catoni, R., & Varone, L. (2014). Morphological, anatomical and physiological leaf traits of *Q. ilex*, *P. latifolia*, *P. lentiscus*, and *M. communis* and their response to Mediterranean climate stress factors. *Botanical Studies*, 55, 35.
<https://pubmed.ncbi.nlm.nih.gov/28510875/>

Gentili, R., Ambrosini, R., Montagnani, C., Caronni, S., & Citterio, S. (2018). Effect of soil pH on the growth, reproductive investment and pollen allergenicity of *Ambrosia artemisiifolia* L. *Frontiers in Plant Science*, 9, 1335.
<https://doi.org/10.3389/fpls.2018.01335>

Gianoli, E., & González-Teuber, M. (2005). Environmental heterogeneity and population differentiation in plasticity to drought in *Convolvulus chilensis*

(Convolvulaceae). *Evolutionary Ecology*, 19(6), 603-613.

<https://doi.org/10.1007/s10682-005-2220-5>

Gliessman, S. R. (2015). *Agroecology: The ecology of sustainable food systems* (3rd ed.). CRC Press.

Gliessman, S. R. (2018). Defining agroecology. *Agroecology and Sustainable Food Systems*, 42(6), 599-600.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gotelli, N. J., & Ellison, A. M. (2013). *A primer of ecological statistics* (2nd ed.). Sinauer Associates.

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2-3), 89-100. [https://doi.org/10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1)

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.

Hänsch, R., & Mendel, R. R. (2009). Physiological functions of mineral micronutrients (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). *Current Opinion in Plant Biology*, 12(3), 259-266. <https://doi.org/10.1016/j.pbi.2009.05.006>

Härdle, W. K., & Simar, L. (2015). *Applied multivariate statistical analysis* (4th ed.). Springer.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall/CRC.

Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10, 4-10.

<https://doi.org/10.1016/j.wace.2015.08.001>

Haz Revista. (2018). Agricultura en huerto vertical Aerofarms [Fotografía]. Haz Revista. <https://hazrevista.org/wp-content/uploads/2018/08/agricultura-huerto-vertical-aerofarms.jpg>

Hobbs, P. R., Sayre, K., & Gupta, R. (2008). The role of conservation agriculture in sustainable agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491), 543-555.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Wadsworth Cengage Learning.

Huang, Y., Wang, J., & Li, F. (2021). Descriptive statistical approaches for analyzing plant growth responses to environmental variables. *Plant Methods*, 17, Article 58.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, 105(2), 302-308.

<https://doi.org/10.1037/0033-2909.105.2.302>

Hunt, R. (1990). *Basic growth analysis: plant growth analysis for beginners*. Springer Science & Business Media.

Hydro Environment. (s.f.). *Cultivo de arándano en invernadero: Guía intensiva para blueberry*. Artículo web, consultado 2023, en: <https://hydroenv.com.mx/cultivo-de-arandano-en-invernadero-guia-intensiva-para-blueberry/>

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Intagri S.C. (2018, enero 27). *El cultivo de arándano o blueberry*. Recuperado el 27 de abril de 2025, de <https://www.intagri.com/articulos/frutillas/El-Cultivo-de-Ar%C3%A1ndano-o-Blueberry>

International Organization for Standardization. (2009). ISO/IEC 20926:2009 - Software and systems engineering — Software measurement — IFPUG functional size measurement method. International Organization for Standardization.

Instituto Colombiano Agropecuario. (2010). Resolución 970 de 2010: Regulación sobre el uso de semillas certificadas. ICA. <https://www.ica.gov.co>

Instituto Colombiano Agropecuario. (2022). Resolución 1056 de 2022: Uso de productos fitosanitarios. ICA. <https://www.ica.gov.co>

Jevšenak, J., & Levanič, T. (2018). dendroTools: R package for studying linear and nonlinear responses between tree-rings and daily environmental data.

Dendrochronologia, 48, 32-39. <https://doi.org/10.1016/j.dendro.2018.01.005>

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson Prentice Hall.

Juroszek, P., & von Tiedemann, A. (2015). Linking plant disease models to climate change scenarios to project future risks of crop diseases: A review. *Journal of Plant Diseases and Protection*, 122(1), 3-15. <https://doi.org/10.1007/BF03356525>

Kimball, B. A. (2016). Crop responses to elevated CO₂ and interactions with H₂O, N, and temperature. *Current Opinion in Plant Biology*, 31, 36-43. <https://doi.org/10.1016/j.pbi.2016.03.006>

Klingenberg, C. P. (2016). Size, shape, and form: Concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, 226(3), 113-137. <https://doi.org/10.1007/s00427-016-0539-2>

Kochian, L. V., Piñeros, M. A., Liu, J., & Magalhaes, J. V. (2015). Plant adaptation to acid soils: The molecular basis for crop aluminum resistance. *Annual Review of Plant Biology*, 66, 571-598. <https://doi.org/10.1146/annurev-arplant-043014-114822>

Lal, R. (2020). Precision agriculture for soil carbon sequestration. *Precision Agriculture*, 21(4), 999-1003.

Lambers, H., Chapin III, F. S., & Pons, T. L. (2008). *Plant physiological ecology* (2nd ed.). Springer.

Larcher, W. (2003). *Physiological plant ecology: ecophysiology and stress physiology of functional groups*. Springer.

Leakey, A. D., Ainsworth, E. A., Bernacchi, C. J., Rogers, A., Long, S. P., & Ort, D. R. (2009). Elevated CO₂ effects on plant carbon, nitrogen, and water relations: Six important lessons from FACE. *Journal of Experimental Botany*, 60(10), 2859-2876. <https://doi.org/10.1093/jxb/erp096>

Leandro, R., Guzmán, F., & Vargas, C. (2019). *Técnicas descriptivas multivariantes: Una guía práctica para su aplicación*. Ediciones UNAM.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. <https://doi.org/10.1038/nature14539>

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons.

Losada Pardo, C. (2021, abril 5). Los arándanos: El nuevo producto con potencial de exportación de Colombia [Los arándanos: el nuevo producto con potencial de exportación de Colombia]. Metroflor. <https://www.metroflorcolombia.com/los-arandanos-el-nuevo-producto-con-potencial-de-exportacionde-colombia/>

Manzano, A., Jiménez, M., & Rodríguez, J. (2019). *Modelos comparativos aplicados a la investigación en ciencias sociales*. Editorial Académica.

Martínez-Herrera, M., Sánchez-García, A., & López-Pérez, R. (2019). Integración de modelos comparativos y descriptivos en el análisis de datos ecológicos. *Revista de Ecología Aplicada*, *18*(1), 71-84.

M. G. KENDALL, A NEW MEASURE OF RANK CORRELATION, *Biometrika*, Volume 30, Issue 1-2, June 1938, Pages 81–93, <https://doi.org/10.1093/biomet/30.1-2.81>

Ministerio de Agricultura, Ganadería y Alimentación de Guatemala (MAGA). (2024). *Ficha técnica agroclimática: Arándano (Vaccinium corymbosum)*. Ciudad de Guatemala: MAGA – DIGEGR.

- Ministerio de Agricultura y Desarrollo Rural. (2015). Decreto 1071 de 2015: Decreto Único Reglamentario del Sector Agropecuario.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=76838>
- Ministerio de Agricultura y Desarrollo Rural. (2019). Resolución 180594 de 2019: Reglamento Técnico de Fertilizantes. <https://www.minagricultura.gov.co>
- Ministerio de Agricultura y Desarrollo Rural. (2021). Política Nacional de Agricultura de Precisión. <https://www.minagricultura.gov.co>
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358-371.
- Ouzounis, T., Rosenqvist, E., & Ottosen, C. O. (2015). Spectral effects of artificial light on plant physiology and secondary metabolism: A review. *HortScience*, 50(8), 1128-1135. <https://doi.org/10.21273/HORTSCI.50.8.1128>
- Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, 7, e6876. <https://doi.org/10.7717/peerj.6876>
- Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., Bret-Harte, M. S., Cornwell, W. K., Craine, J. M., Gurvich, D. E., Urcelay, C., Veneklaas, E. J., Reich, P. B., Poorter, L., Wright, I. J., Ray, P., Enrico, L., Pausas, J. G., de Vos, A. C., ... Cornelissen, J. H. C. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, 61(3), 167-234. <https://doi.org/10.1071/BT12225>

Pérez-López, E., & Rodríguez-Gómez, M. (2019). Técnicas de análisis descriptivo para la caracterización del crecimiento vegetal en estudios ecofisiológicos. *Ecosistemas*, 28(3), 456-470.

Pierce, F. J., & Nowak, P. (1999). Aspects of precision agriculture. *Advances in Agronomy*, 67, 1-85.

Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences*, 109(31), 12302-12308.

Polo, V. 2013. Modelos matemáticos en ecología: aplicación al dilema halcón vs. paloma. *Ecosistemas* 22(3):6-11. 10.7818/ECOS.2013.22-3.02

Poorter, H., & Nagel, O. (2000). The role of biomass allocation in the growth response of plants to different levels of light, CO₂, nutrients and water: A quantitative review. *Australian Journal of Plant Physiology*, 27(6), 595-607.
<https://doi.org/10.1071/PP99173>

Poorter, H., Niinemets, Ü., Poorter, L., Wright, I. J., & Villar, R. (2012). Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytologist*, 193(1), 30-50. [10.1111/j.1469-8137.2009.02830.x](https://doi.org/10.1111/j.1469-8137.2009.02830.x)

Poorter, H., Niinemets, Ü., Ntagkas, N., Siebenkäs, A., Mäenpää, M., Matsubara, S., & Pons, T. L. (2019). A meta-analysis of plant responses to light intensity for 70 traits ranging from molecules to whole plant performance. *New Phytologist*, 223(3), 1073-1105. <https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15754>

Poorter, H., Niklas, K. J., Reich, P. B., Oleksyn, J., Poot, P., & Mommer, L. (2012). Biomass allocation to leaves, stems and roots: Meta-analyses of interspecific variation

and environmental control. *New Phytologist*, 193(1), 30-50.

<https://doi.org/10.1111/j.1469-8137.2011.03952.x>

Prasad, P. V. V., Bheemanahalli, R., & Jagadish, S. V. K. (2017). Field crops and the fear of heat stress—Opportunities, challenges and future directions. *Field Crops Research*, 200, 114-121. <https://doi.org/10.1016/j.fcr.2016.09.024>

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press.

Rankel, K. (2024, 18 junio). *How Much Humidity Does My American Blueberry Need?* Greg Blog. Recuperado de <https://greg.app/american-blueberry-humidity/>

Reich, P. B., Wright, I. J., Cavender-Bares, J., Craine, J. M., Oleksyn, J., Westoby, M., & Walters, M. B. (2003). The evolution of plant functional variation: Traits, spectra, and strategies. *International Journal of Plant Sciences*, 164(S3), S143-S164. <https://doi.org/10.1086/374368>

Reganold, J. P., & Wachter, J. M. (2016). Organic agriculture in the twenty-first century. *Nature Plants*, 2(2), 15221.

Rietra, R. P., Heinen, M., Dimkpa, C. O., & Bindraban, P. S. (2017). Effects of nutrient antagonism and synergism on yield and fertilizer use efficiency. *Communications in Soil Science and Plant Analysis*, 48(16), 1895-1920. <https://doi.org/10.1080/00103624.2017.1407429>

Robert, P. C. (2002). Precision agriculture: A challenge for crop nutrition management. *Plant and Soil*, 247(1), 143-149.

Rousk, J., Brookes, P. C., & Bååth, E. (2009). Contrasting soil pH effects on fungal and bacterial growth suggest functional redundancy in carbon mineralization. *Applied and Environmental Microbiology*, 75(6), 1589-1596.

<https://doi.org/10.1128/AEM.02775-08>

Rüger, N., Wirth, C., Wright, S. J., & Condit, R. (2012). Functional traits explain light and size response of growth rates in tropical tree species. *Ecology*, 93(12), 2626-2636. <https://doi.org/10.1890/12-0622.1>.

Sadler, E. J., Evans, R. G., Stone, K. C., & Camp, C. R. (2005). Opportunities for conservation with precision irrigation. *Journal of Soil and Water Conservation*, 60(6), 371-378.

Stafford, J. V. (2000). Implementing precision agriculture in the 21st century. *Journal of Agricultural Engineering Research*, 76(3), 267-275.

Seufert, V., Ramankutty, N., & Foley, J. A. (2012). Comparing the yields of organic and conventional agriculture. *Nature*, 485(7397), 229-232.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177. <https://doi.org/10.1037/1082-989X.7.2.147>

Scheiner, S. M. (2001). MANOVA: Multiple response variables and multispecies interactions. In S. M. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (2nd ed., pp. 99-115). Oxford University Press.

Scheiner, S. M., & Gurevitch, J. (Eds.). (2001). *Design and analysis of ecological experiments* (2nd ed.). Oxford University Press.

Smith, J. R., & Johnson, A. B. (2022). Descriptive statistics in plant growth analysis: A comprehensive guide. En R. Brown & S. Davis (Eds.), *Modern approaches in plant science research* (pp. 112-135). Academic Press.

Solinftec. (2022). *Agricultura de precisión [Fotografía]*. Solinftec. <https://www.solinftec.com/wp-content/uploads/2022/07/Agricultura-de-precisio%CC%81n-1080x675.jpeg>

Song, Y. H., Shim, J. S., Kinmonth-Schultz, H. A., & Imaizumi, T. (2015). Photoperiodic flowering: Time measurement mechanisms in leaves. *Annual Review of Plant Biology*, 66, 441-464. <https://doi.org/10.1146/annurev-arplant-043014-115555>

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Tardieu, F., Simonneau, T., & Muller, B. (2018). The physiological basis of drought tolerance in crop plants: A scenario-dependent probabilistic approach. *Annual Review of Plant Biology*, 69, 733-759. <https://doi.org/10.1146/annurev-arplant-042817-040218>

Tamayo Urgilés, D. A. (2023). *Construcción de un dataset de eventos de conducción utilizando modelos de generación de datos sintéticos mediante Generative Adversarial Networks (GAN)* [Tesis de maestría, Escuela Politécnica Nacional]. Repositorio Digital EPN. <https://bibdigital.epn.edu.ec/handle/15000/24766>

Varela-Arregoces, E., & Campbells-S, E. (2011). *Redes Neuronales Artificiales: una revisión del estado del arte, aplicaciones y tendencias futuras*. Universidad Simón Bolívar. Investigación y Desarrollo en TIC.

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman and Hall/CRC.

von Ende, C. N. (2001). Repeated-measures analysis: Growth and other time-dependent measures. In S. M. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (2nd ed., pp. 134-157). Oxford University Press.

Way, D. A., & Yamori, W. (2014). Thermal acclimation of photosynthesis: On the importance of adjusting our definitions and accounting for thermal acclimation of respiration. *Photosynthesis Research*, 119(1), 89-100. <https://doi.org/10.1007/s11120-013-9873-7>

Wei, W. A. S. (2006). *Time series analysis: univariate and multivariate methods* (2nd ed.). Pearson.

White, P. J., George, T. S., Gregory, P. J., Bengough, A. G., Hallett, P. D., & McKenzie, B. M. (2013). Matching roots to their environment. *Annals of Botany*, 112(2), 207-222. <https://doi.org/10.1093/aob/mcs118>

Winsor, C. P. (1932). The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences*, 18(1), 1-8. <https://doi.org/10.1073/pnas.18.1.1>

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>

Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, J. H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, *91*(3), 361-371.

<https://doi.org/10.1093/aob/mcg029>

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture*, *36*(2-3), 113-132.

Zhang, X., Liu, Y., & Chen, H. (2023). Integrating descriptive analysis techniques with environmental data to assess plant growth patterns. *Journal of Plant Ecology*, *16*(2), 245-260.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer.

<https://doi.org/10.1007/978-0-387-87458-6>

Zuur, A. F., Ieno, E. N., & Smith, G. M. (2007). *Analysing ecological data*. Springer Science & Business Media. <https://doi.org/10.1007/978-0-387-45972-1>

4. Anexos



UDECA
UNIVERSIDAD DE
CUNDINAMARCA

**Manual/Guía de Usuario/Técnico
Desarrollo de Sistemas de Información
Dirección de Sistemas y Tecnología**

TABLA DE CONTENIDO

1	INTRODUCCIÓN.....	3
1.1	CONTEXTO DEL PROYECTO	3
1.2	ALCANCE DEL SISTEMA.....	3
1.3	REQUISITOS PREVIOS	3
2	ACCESO AL SISTEMA	3
3	FUNCIONALIDADES DEL SISTEMA.....	5
3.1	INTERFAZ WEB	5
3.2	VISUALIZACIÓN DE DATASETS	5
3.3	VISUALIZACIÓN DE METODOLOGÍA.....	8
3.4	VISUALIZACIÓN DEL MODELO MATEMÁTICO.....	11
3.5	GENERACIÓN DE REPORTES.....	14
4	ROLES Y PERMISOS	15
5	SOLUCIÓN DE PROBLEMAS	15
6	DATOS DE CONTACTO SOPORTE APLICACIÓN.....	17
7	DIRECTORES DE SISTEMAS Y TECNOLOGÍA	17

1 INTRODUCCIÓN

Este sistema fue desarrollado con base al proyecto macro “Prototipo de Sistema Computacional para Gestionar y Controlar un Cultivo de Bayas Silvestres en un Sistema Cerrado de Recirculación de Agua en la Universidad de Cundinamarca” para modelar y analizar el crecimiento de la planta de arándano Biloxi en sistemas abiertos y cerrados, considerando variables ambientales clave como temperatura, humedad, luminosidad y pH. Su objetivo es brindar una herramienta tecnológica que facilite la investigación agrícola y la toma de decisiones basada en datos, contribuyendo a la agricultura de precisión

1.1 CONTEXTO DEL PROYECTO

- **Importancia del cultivo de arándanos Biloxi:** Esta variedad es reconocida por su adaptabilidad y valor comercial, pero su crecimiento depende críticamente de condiciones ambientales controladas.
- **Innovación tecnológica:** El software integra modelos matemáticos con datos reales y sintéticos, permitiendo predecir tener un análisis mas profundo sobre el comportamiento de cada una de las variables además de su impacto en el crecimiento.

1.2 ALCANCE DEL SISTEMA

- Visualización de datasets en tiempo real.
- Generación de reportes personalizados en PDF.
- Acceso a modelos matemáticos y comprensión de estos resultados por medio de información.

1.3 REQUISITOS PREVIOS

Técnicos:

- Navegador web actualizado (Chrome, Firefox o Edge).
- Conexión a internet estable.

De usuario:

- **Usuario estándar:** Registro básico en el sistema.
- **SuperAdministrador:** Credenciales de acceso privilegiado por otro **SuperAdministrador**.

2 ACCESO AL SISTEMA

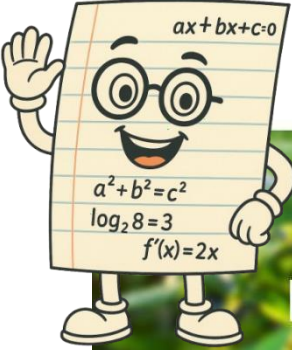
El inicio de sesión no es una parte propia de este documento, al ser este un módulo el acceso viene de la aplicación principal.

URL de acceso: Proporcionar el enlace web del sistema.

Inicio de sesión:

- Paso 1: Ingresar nombre de usuario y contraseña.
- Paso 2: Hacer clic en "Iniciar Sesión".

Nota: Los **SuperAdministradores** tendrán acceso a funciones adicionales.




$ax+bx+c=0$
 $a^2+b^2=c^2$
 $\log_2 8=3$
 $f'(x)=2x$

Iniciar Sesión **Registrarse**

Selecciona el botón de Iniciar Sesión

Bienvenido a B Berry Company

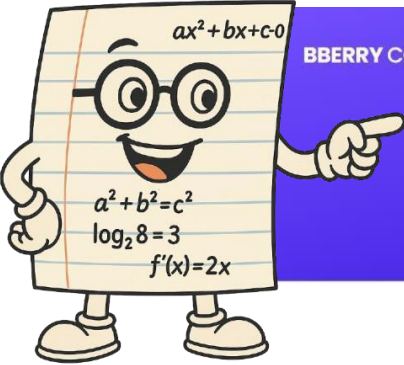
¡Aquí conocerás y descubrirás sobre el arándano y su cuidado en tus cultivos!



Seguimiento de cultivos

Control de variables

Reportes 24/7



$ax^2+bx+c=0$
 $a^2+b^2=c^2$
 $\log_2 8=3$
 $f'(x)=2x$

BBERRY COMPANY

Inicia Sesión

Ingresar tu email y contraseña para iniciar sesión!

Email

Contraseña
Min. 8 caracteres incluido al menos un caracter especial

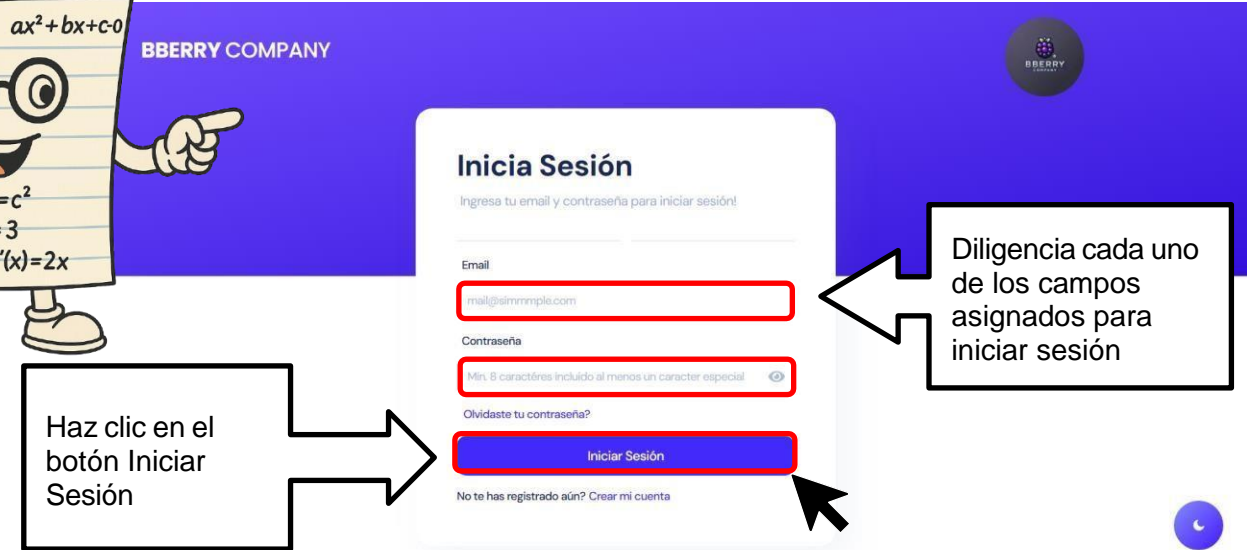
Olvidaste tu contraseña?

Iniciar Sesión

No te has registrado aún? [Crear mi cuenta](#)

Haz clic en el botón Iniciar Sesión

Diligencia cada uno de los campos asignados para iniciar sesión



3 FUNCIONALIDADES DEL SISTEMA

3.1 INTERFAZ WEB

Objetivo: Navegar por las vistas principales del sistema.

Pasos:

- Tras iniciar sesión, se mostrará la página principal con un menú lateral de opciones:
 - Datasets.
 - Ambientes.
 - Reportes.



Hacer clic en cualquier opción para acceder a la vista correspondiente.

3.2 VISUALIZACIÓN DE DATASETS

Objetivo: Consultar y descargar los datos utilizados en el modelo.

Pasos:

- En el menú lateral, seleccione "Datasets".
 - Se mostrará una vista de varias tablas con los datos ambientales y de crecimiento.
- Para descargar:
- Hacer clic en "Descargar CSV" o "Descargar" el dataset completo.

Pages / Datasets

Datasets

Q Buscar...

Dataset Completo

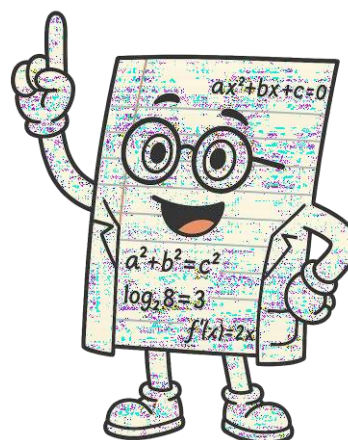
Q Buscar...

Realice búsquedas específicas de todos los datasets

Visualice tablas con información correspondiente de los datasets

			LUM	TEMP	ALTURA
2024-01-01 01:00:00	91.615	100	311.148	17.598	14.333
2024-01-01 02:00:00	91.485	100	297.099	17.324	14.333
2024-01-01 03:00:00	91.703	100	264.947	17.184	14.333
2024-01-01 04:00:00	92.419	100	321.24	17.212	14.333
2024-01-01 05:00:00	93.133	100	452.794	17.229	14.333
2024-01-01 06:00:00	93.822	100	599.833	17.212	14.333
2024-01-01 07:00:00	93.545	100	713.028	17.187	14.333

Haz clic en los botones Descargar CSV para obtener el dataset que se desee



Efecto parcial Temperatura

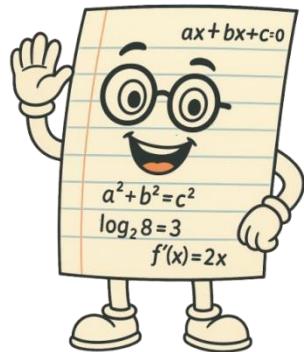
Q Buscar...

TEMP	PARTIAL EFFECT	CONFIDENCE LOWER	CONFIDENC
15.102	1.545	0.878	2.212
15.182	1.546	0.944	2.149
15.261	1.552	1.01	2.095
15.341	1.563	1.076	2.049
15.421	1.576	1.141	2.012
15.501	1.593	1.203	1.983
15.58	1.611	1.259	1.963
15.66	1.631	1.309	1.952

Efecto parcial H. Sustrato

Q Buscar...

HUM SUSTR	PARTIAL EFFECT	CONFIDENCE LOWER	CONFIDENC
60.713	2.023	1.581	2.465
61.11	2.031	1.643	2.418
61.507	2.032	1.69	2.374
61.904	2.028	1.722	2.335
62.3	2.02	1.739	2.301
62.697	2.007	1.74	2.274
63.094	1.99	1.727	2.254
63.491	1.971	1.703	2.24



CAS DE FERTILIZ... DATASHEETS PROYECTO ELECTRONICA YouTube VirusTotal - Home Programas básicos...

Pages / Datasets

Datasets

Archivos disponibles

Q Buscar...

ARCHIVO	DESCARGAR
Modelo Gam Semicontrolado 17-03-2025 (.pkl)	Descargar
Dataset Completo	Descargar
Efecto parcial Humedad Relativa	Descargar
Efecto parcial Humedad Sustrato	Descargar
Efecto parcial Luminosidad	Descargar
Efecto parcial Minutos Siembra	Descargar
Efecto parcial Poda Engrosamiento	Descargar
Efecto parcial Temperatura Alta	Descargar

Página 1 de 2

1. Seleccione el dataset que desee descargar en su máquina local

2. Validación de dataset descargado, haz clic para mirar dataset

3.3 VISUALIZACIÓN DE METODOLOGÍA

Objetivo: Revisar el proceso de construcción del modelo.

Pasos:

- En el menú lateral principal, seleccionar "**Reporte Modelado**".

Se desplegará una vista con:

- Metodología y modelos
- Origen de los datos.
- Técnicas de procesamiento.

Visualice información en metodologías y técnicas usadas para la elaboración del modelo



Metodología y Modelos

Origen y Preprocesamiento de los Datos

Ambientes del prototipo

El **sistema** cuenta con tres ambientes diferenciados por su grado de control:

- **Semicontrolado:** 3 plantas y 2 grupos de sensores (capacitivo de humedad, fotorresistencia, DHT22 y sensor UV). Combina luz artificial con luz solar y flujo de aire exterior.
- **Controlado:** 1 planta y 1 grupo de sensores. Recibe muy poca luz solar y está mayormente cerrado, con flujo de aire limitado.
- **Abierto:** exposición casi total a la luz solar, sin control de ventilación.

Datos utilizados

Para este estudio se emplearon únicamente los datos del ambiente semicontrolado: las lecturas de los dos grupos de sensores y la altura de las 3 plantas.

Selección del modelo para generar datos sintéticos

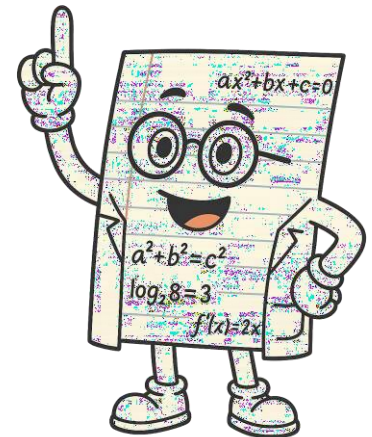
Dada la forma no lineal de la relación entre variables ambientales y altura, se optó por una **red neuronal**, capaz de capturar patrones complejos.

Arquitectura de la red neuronal

- 3 capas ocultas con 64, 32 y 16 neuronas respectivamente.
- Función de activación ReLU en cada capa oculta.
- Una neurona de salida para regresión del dato a predecir.

Entrenamiento

- **Épocas:** 300, para un ajuste más fino a los datos reales.
- Optimización mediante Adam y función de pérdida MSE.



Modelo Aditivo Generalizado

GAM-LinearGAM-pyGAM

Modelo GAM

$$y_i = B_0 + \sum_{j=1}^p \left(\sum_{k=1}^{K_j} B_{jk} b_{jk}(x_{ij}) \right)$$

- **Intercepto (B_0):** Constante del modelo
- **Predictores (p):** Número total de variables explicativas
- **Splines por predictor (K_j):** Grados de libertad de cada spline
 - **Coefficientes (B_{jk}):** Peso de cada función base
 - **Funciones base ($b_{jk}(x_{ij})$):** Spline evaluado en x_{ij}

Desarrollo e Implementación del Modelo GAM

Enfoque no lineal y penalización

Este enfoque captura relaciones no lineales complejas sin imponer una forma rígida. Se emplea cuando se requiere más flexibilidad que en la regresión lineal tradicional, controlando el sobreajuste mediante penalizaciones del hiperparámetro Lambda.

Herramientas y librerías

El modelo se desarrolló en **Python** usando:

- pandas y numpy para manejo de datos.
- pygam. LinearGAM para construir el GAM con splines ($s()$) y términos categóricos ($f()$).
- sklearn.metrics (r2_score, mean_squared_error).

Variables predictoras y objetivo

Las variables predictoras provienen de sensores ambientales (temperatura, humedad, luz, etc.), acciones de siembra, intensidad de poda, etc.). La variable objetivo es la curva de influencia de cada predictor en el GAM. [Ver dataset general](#)

Redirecciones dinámicas a otras secciones del módulo de Reportes

Más información de GAM

Los GAM extienden la regresión lineal permitiendo sumar funciones suaves de cada predictor. Cada término $s(i)$ es un spline penalizado que capta no linealidades, y $f(j)$ modela variables categóricas. La suavidad se controla con el parámetro λ , equilibrando flexibilidad y sobreajuste.

R² (Coeficiente de determinación)

Mide qué proporción de la variabilidad de los datos reales explican las predicciones. Valores cercanos a 1 indican un ajuste más preciso.

RMSE (Root Mean Square Error)

Error cuadrático medio: cuanto menor, mejor. Representa la desviación promedio entre predicciones y observaciones reales.

Distribución de errores y función de enlace

Se asume distribución normal de los residuos. La función de enlace conecta la suma de splines con la escala de la variable objetivo (identity, log, etc.).

Número de muestras

Cantidad de observaciones empleadas para ajustar el modelo. Más datos suelen estabilizar la estimación de splines.

f(j) – Término categórico

Función lineal para variables categóricas (factoriales). p.ej. $f(7)$ indica un dummy con 2 grados de libertad.

λ , Rank y EDoF

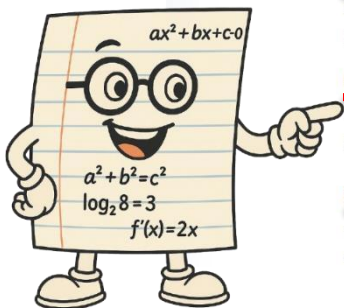
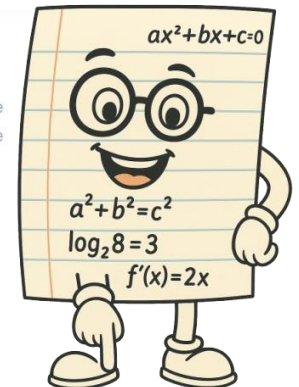
• λ : parámetro de penalización de cada spline. • Rank/EDoF: grados de libertad efectivos del spline; mayor EDoF \Rightarrow curva más compleja.

P > x y Sig Code

– P > x: p-valor de la prueba nula. – Sig Code: asteriscos indican importancia (***) muy significativo; . marginal; sin asteriscos no significativo).

Más sobre GAM

- El ajuste se realiza vía algoritmo de backfitting y optimización de λ por GCV/AIC.
- Permite interpretar cada efecto parcial en gráficos de dependencia parcial.



3.4 VISUALIZACIÓN DEL MODELO MATEMÁTICO

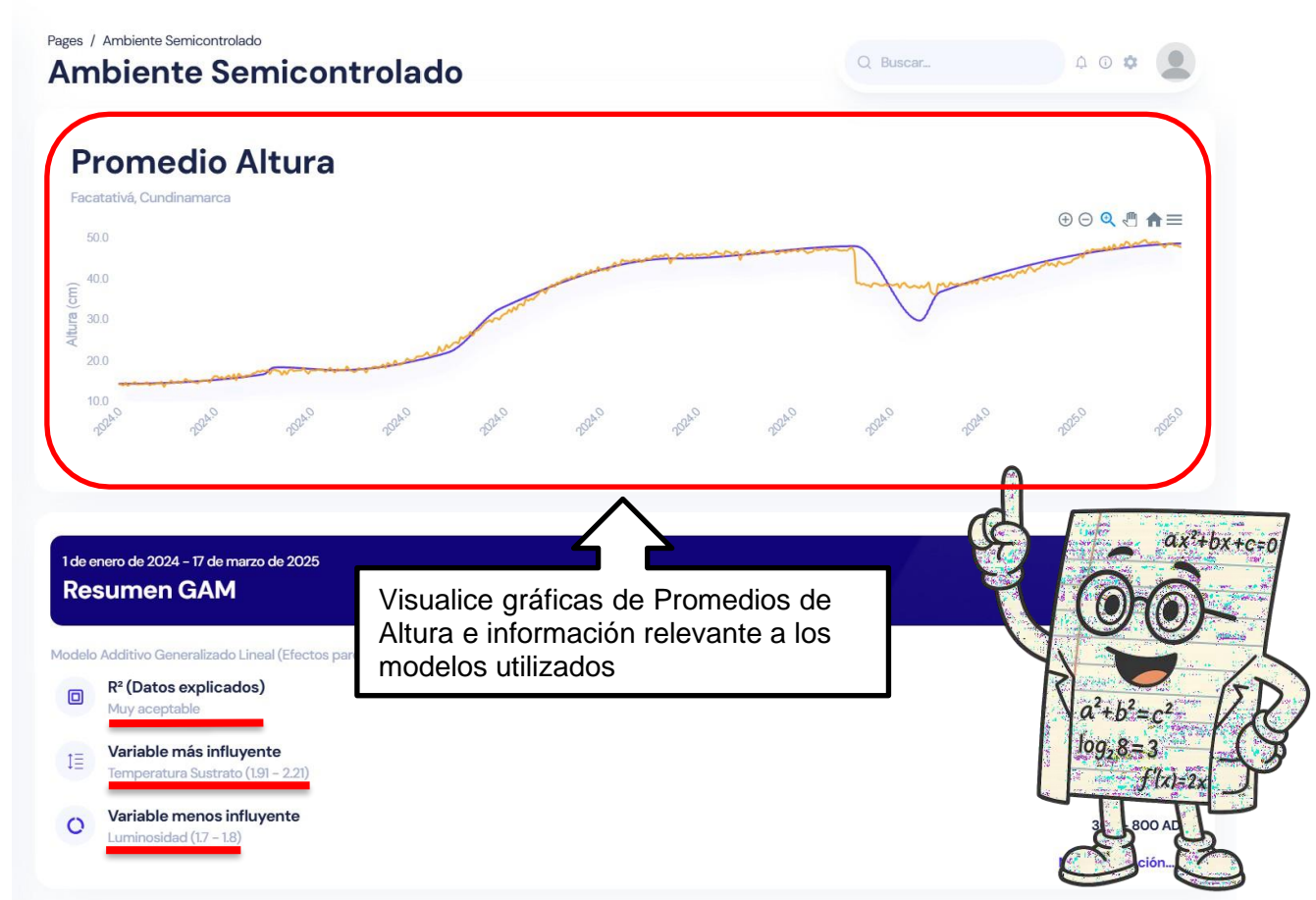
Objetivo: Analizar el modelo y sus variables.

Pasos:

- En el menú lateral principal, seleccionar el **Ambiente** al que se desea conocer como interactúa el modelo matemático con dicho Ambiente.

En la pantalla se mostrará una interfaz con lo siguiente:

- Ecuaciones del modelo.
- Gráficas lineales de correlación entre variables ambientales y crecimiento.
- Parámetros estadísticos (coeficiente de correlación).



Importante, las demás vistas de los Ambientes cuentan con la misma estructura de información en interfaz de visualización solo difieren en contenido y en la forma de las gráficas.

Variables y efectos parciales

Información del modelo

Logística y Gompertz

Visualice información respecto a ecuaciones utilizados en el modelo

Modelo Gompertz
2 de noviembre de 2024

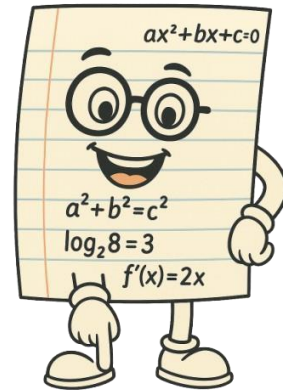
$$y(x) = 16.283 + 31.106 \cdot \exp(-\exp(-0.032(x - 147.846)))$$

- Nivel Base: 16.283 cm
- Inflexión: 147.846 días
- Tasa de crecimiento: -0.032
- Asíntota Superior: 47.389 cm (16.283 + 31.106)

Modelo Logístico
2 de noviembre de 2024

$$y(x) = 15.534 + \frac{31.247}{1 + e^{-0.045(x - 158.934)}}$$

- Nivel Base: 15.534 cm
- Inflexión: 158.934 días
- Tasa de crecimiento: -0.045
- Altura Adicional: 46.781 cm (15.534 + 31.247)

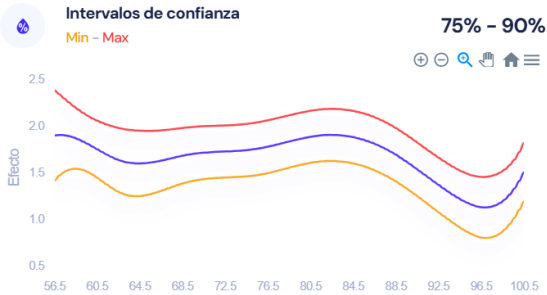


Variables GAM

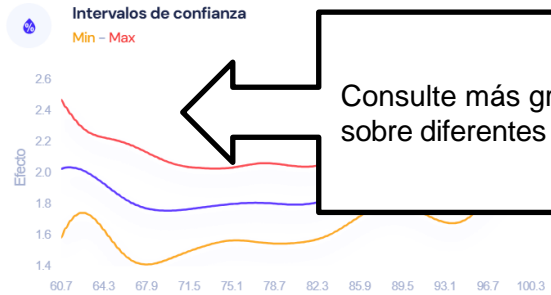
Ver efectos

NOMBRE	GRADOS LIBERTAD	GRADOS LIBERTAD EFECTIVOS	P > x	SIG. CODE
Temperatura alta	10	8.8	1.11e-16	***
Temperatura mínima	10	8.5	1.11e-16	***
Humedad relativa en aire	10	7.6	1.11e-16	***
Tiempo desde la siembra	10	8.5	1.11e-16	***
Humedad de sustrato	10	7.8	6.66e-16	***
Luminosidad	10	7.6	7.35e-01	
Temperatura	10	7.0	3.43e-04	***
Poda Engrosamiento	2	1.0	1.11e-16	***

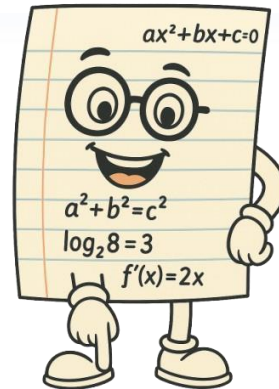
Humedad Relativa en Aire



Humedad Sustrato



Consulte más gráficas sobre diferentes variables

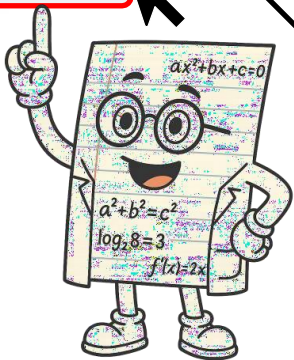


Visualice más gráficas de las diferentes variables monitoreadas



Haz clic en el botón Descargar PDF para obtener un reporte completo de lo visualizado

Descargar PDF



3.5 GENERACIÓN DE REPORTE

Objetivo: Descargar reportes en PDF.

Pasos:

- En cualquier vista (**Ambiente Controlado, Ambiente Semicontrolado, Ambiente No Controlado**)
- Hacer clic en "**Descargar PDF**".

TIPOS DE SUELOS TIPOS DE FERTILIZANTE TECNICAS DE FERTILIZ... DATASHEETS PROYECTO ELECTRONICA YouTube VirusTotal - Home Programas básicos...

BBERRY COMPANY

Páginas / Ambiente Semicontrolado

Ambiente Semicontrolado

Inicio Administración Ambientes Reportes Reporte Modelado Ambiente Controlado Ambiente Semicontrolado Ambiente No Controlado Datasets

Temperatura Alta Intervalos de confianza 18° - 23°

Temperatura Min Intervalos

Descargar PDF

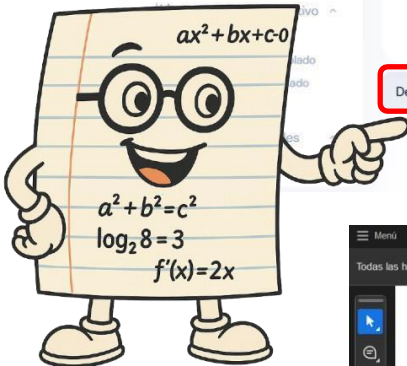
1. Haz clic en el botón Descargar PDF

Historial de descargas recientes

page.pdf 4.7 MB - Hecho

Historial de descargas completo

2. Haz clic en el PDF descargado



Menú page.pdf + Crear Iniciar sesión

Todas las herramientas Editar Convertir Firma electrónica Buscar texto o herramientas

Promedio Altura

Facultad de Ingeniería

Altura (cm)

1 de enero de 2024 - 17 de marzo de 2025

Resumen GAM

Modelo Aditivos Generalizado Lineal (3 Factores parciales)

R² (Datos explicados) 97.83%
Muy aceptable

Variable más influyente 85% - 100%
Temperatura Ambiente (1.01 - 2.21)

Variable menos influyente 300 - 800 ADC
Luminosidad (1.7 - 1.8)

Variables y efectos parciales

Información del modelo

Log Preguntas al Asistente de IA ¿Por qué es importante esta información?

PDF descargado y visualizado en la máquina local

4 ROLES Y PERMISOS

Usuario:

Acceso a visualización de datos, modelos y generación de reportes.

SuperAdministrador:

Funciones adicionales:

- Configuración de microservicios.
- Actualización de datasets.
- Gestión de usuarios.

5 SOLUCIÓN DE PROBLEMAS

Esta sección aborda los problemas más comunes que los usuarios pueden enfrentar al interactuar con el sistema, junto con soluciones prácticas basadas en los requisitos técnicos y funcionales descritos en el documento original.

Error: problemas de acceso y autenticación "Credenciales incorrectas"

Causas:

- Usuario no registrado en el sistema.
- Contraseña mal escrita o caducada.

Solución: Verificar que el nombre de usuario y contraseña sean correctos.

Si es **SuperAdministrador**, asegurarse de usar credenciales con permisos elevados.

Contactar al soporte técnico para restablecer la contraseña si es necesario.

Error: "Acceso no autorizado a microservicios"

Causa:

- Token de autenticación inválido o expirado (según RNF-02).

Solución: Cerrar sesión y volver a iniciar para regenerar el token.

Asegurarse de que la conexión sea mediante HTTPS (protocolo seguro).

Problemas con Visualización de Datos: Gráficas o tablas no se cargan

Causas:

- Navegador no compatible con HTML5 o JavaScript (según restricciones de diseño).
- Fallas en la conexión a la base de datos SQL Server (entorno operativo).

Solución: Usar navegadores actualizados como Chrome, Firefox o Edge, limpiar la caché del navegador, verificar la conexión a internet.

Problemas con Generación de Reportes: Error al descargar reportes en PDF

Causas:

- Bloqueo de ventanas emergentes en el navegador.
- Espacio insuficiente en el dispositivo.

Solución: Desactivar bloqueadores de ventanas emergentes, liberar espacio en el disco duro e intentar generar el reporte nuevamente desde la opción "Descargar PDF".

Datos inconsistentes en el modelo matemático

Causa:

- Datasets no actualizados o errores en la integración con Python.

Solución:

Los **SuperAdministradores** deben verificar la sincronización de datos en los formatos correctos CSV/JSON, revisar el código del modelo en Python para ajustar parámetros.

Problemas de Rendimiento: Lentitud en el sistema con múltiples usuarios

Causa:

- Límite de concurrencia (máximo 5 usuarios simultáneos).

Solución:

Reducir el número de usuarios activos y los **SuperAdministradores** pueden optimizar la base de datos SQL Server o migrar a bases de datos no relacionales para escalabilidad.

Contactar Soporte Técnico

Si los problemas persisten, proporcione la siguiente información al equipo de soporte:

- Captura de pantalla del error.
- Navegador y versión utilizada.
- Pasos realizados antes del error.
- Rol de usuario (**Usuario** o **SuperAdministrador**).

Correo de soporte: company.bberrry@gmail.com | Teléfono: No Aplica

6 DATOS DE CONTACTO SOPORTE APLICACIÓN

Extensiones telefónicas: No Aplica
Correo: company.bberr@gmail.com

DIRECCIÓN DE SISTEMAS Y TECNOLOGÍA

7 DIRECTORES DE SISTEMAS Y TECNOLOGÍA

Ariel Camilo Sánchez López
Cargo: Pregrado Ingeniería de Sistemas y Computación

Jhonatan Murcia Posso
Cargo: Pregrado Ingeniería de Sistemas y Computación

ÁREA DE DESARROLLO DE SISTEMAS DE INFORMACIÓN



UDEC
UNIVERSIDAD DE
CUNDINAMARCA

CONTROL DE CAMBIOS				
VERSIÓN	FECHA DE APROBACIÓN			DESCRIPCIÓN DEL CAMBIO
	AAAA	MM	DD	
1	2015	04	29	Emisión del documento.
2	2016	05	11	Cambio de portada, página legal, estilo de la letra y contraportada.
3	2016	10	14	Se adiciona en la portada pág. #1 "Manual de Usuario" y el año y se elimina el campo "Oficina de Soporte a Aplicaciones" de la pág. #2
4	2018	07	27	Cambio de portada, estilo de letra, se cambia la "Introducción" por "Objetivo o propósito de la aplicación". Se retira los puntos de "Iconos generales" y "Control de cambios del manual"
5	2019	02	08	Se agregaron los integrantes de la dirección de sistemas y tecnología
6	2020	10	19	Se agregó el ítem de soporte a la aplicación y se realizaron modificaciones para cumplir con la Resolución 000158 del 29 de octubre de 2019, "POR LA CUAL SE ADOPTA LA IMAGEN CORPORATIVA Y SE DICTAN LAS POLÍTICAS DE IDENTIDAD CORPORATIVA DE LA UNIVERSIDAD DE CUNDINAMARCA"
7	2023	01	31	Se realizó cambio del diseño de la portada del documento. Se ajusta proceso como manual/guía para llevar a cabo el formato de estos documentos según la imagen institucional.
8	2024	08	02	Se realiza el cambio en el nombre del formato, incluyendo "Manual/Técnico de Software". Según la resolución rectoral 074 del 22 de julio del 2024, dando cumplimiento a la ley 2345 del 2023 "chao marcas" y dando alcance a la circular 006 "Cambio de identificador visual en los documentos de gestión documental" se realiza el cambio de logo en el documento.
ELABORÓ				
NOMBRES Y APELLIDOS				CARGO
Yohan Sebastián Moreno Lesmes				Técnico IV
María Alejandra Guerrero Martínez				Técnico IV
Yamid Santiago Montañez Monroy				Técnico

REVISO				
NOMBRES Y APELLIDOS		CARGO		
Ana Lucia Hurtado Mesa		Directora de Sistemas y Tecnología		
APROBÓ (GESTOR RESPONSABLE DEL PROCESO)				
NOMBRES Y APELLIDOS	CARGO	FECHA		
		AAAA	MM	DD
Ana Lucia Hurtado Mesa	Directora de Sistemas y Tecnología	2024	08	02



**Manual/Guía de Usuario/Técnico
Desarrollo de Sistemas de Información
Dirección de Sistemas y Tecnología**

TABLA DE CONTENIDO

1	Introducción	3
2	Frontend (Firebase Hosting)	4
2.1	Configuración despliegue frontend	4
2.1.1	Crear proyecto en Firebase Hosting	4
2.1.2	Configuración firebase primer despliegue	5
2.2	Crear archivo de despliegue	6
3	Codificación Modelos Matemáticos	8
3.1	ModeloGam.py	9
3.1.1	Explicación por Fracciones del Código	11
3.1.2	Resultados Obtenidos	15
3.2	ModeloGompertz.py.....	15
3.2.1	Explicación por Fracciones de Código	17
3.2.2	Resultados.....	19
3.3	ModeloLogistico.py	20
3.3.1	Explicación por Fracciones de Código	23
3.3.2	Resultados.....	26
3.4	CorrelacionKendall.py	27
3.4.1	Explicación por Fracciones de Código	27
3.4.2	Resultados.....	28
	DATOS DE CONTACTO SOPORTE APLICACIÓN	30

1 Introducción

El presente documento, constituye una guía integral para la implementación, despliegue y operación de un sistema de información diseñado para mejorar procesos en ambientes controlados, con énfasis en agricultura de precisión. Su estructura técnica aborda tres pilares fundamentales: el despliegue automatizado del frontend, la implementación de modelos matemáticos para análisis predictivo y la configuración segura de bases de datos, integrando herramientas modernas de desarrollo, cloud computing y ciencia de datos.

Componentes clave del manual:

- Despliegue Frontend con Firebase Hosting:
- Instrucciones detalladas para configurar CI/CD mediante GitHub Actions, garantizando despliegues automáticos en Firebase Hosting.
- Gestión de variables de entorno y seguridad con Secrets de GitHub, esencial para entornos productivos.

Modelos Matemáticos en Python:

- Implementación de cuatro modelos analíticos:
 - GAM (Modelo Aditivo Generalizado): Para relaciones no lineales entre variables ambientales y crecimiento vegetal.
 - Gompertz y Logístico: Modelos sigmoidales para predecir curvas de crecimiento en plantas.
 - Correlación de Kendall: Análisis no paramétrico para identificar dependencias entre variables.
- Código completo, explicaciones paso a paso y visualización de resultados, respaldados por librerías como pygam, scipy y seaborn.

Configuración de MongoDB Atlas:

- Creación de clústeres en la nube, gestión de IPs permitidas y generación de cadenas de conexión seguras.
- Buenas prácticas para integración con backend y protección de datos sensibles.

Estructura y utilidad

El manual está organizado en secciones modulares, con capturas de pantalla, diagramas y tablas que facilitan la comprensión técnica. Cada procedimiento incluye:

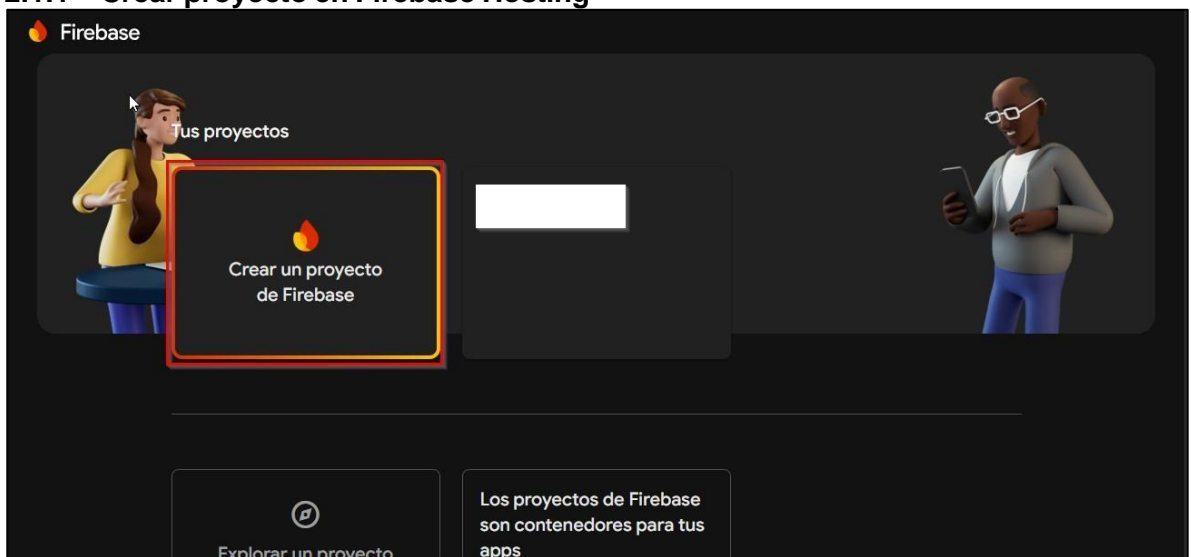
- Contexto teórico: Breve fundamentación de herramientas y metodologías.
- Pasos ejecutables: Comandos, scripts y configuraciones listas para replicar.
- Validación de resultados: Métricas de evaluación (R^2 , RMSE, MSE) y gráficos explicativos.

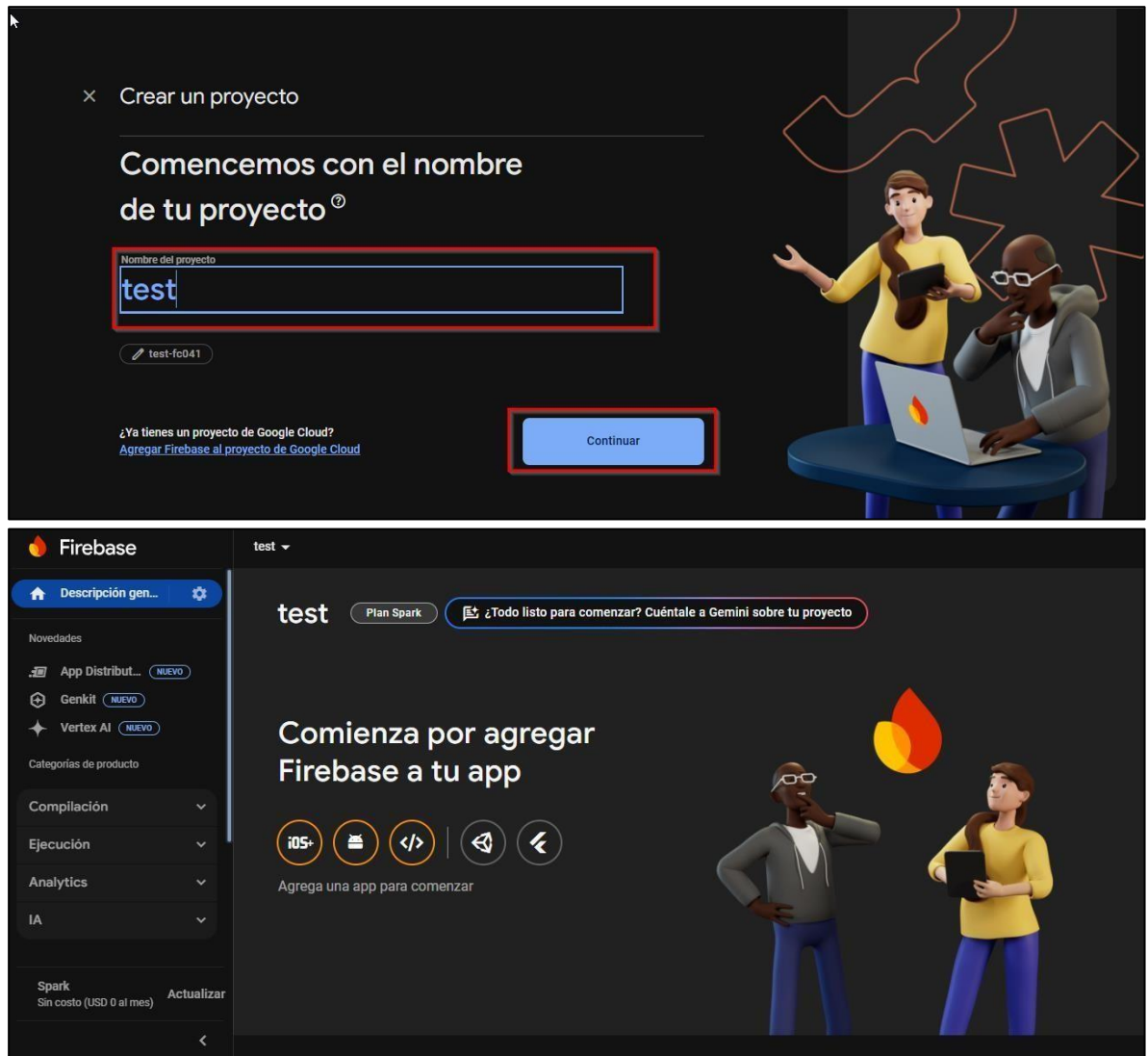
2 Frontend (Firebase Hosting)

El frontend (SPA en React/Vite) se despliega en Firebase Hosting mediante GitHub Actions. Se configura un *release channel* (live) y variables de entorno como `VITE_APP_API_URL`, vinculada al dominio del backend el objetivo es facilitar la implementación continua (CI/CD) de cambios en el código, asegurando que cada actualización en la rama principal (main) se despliegue automáticamente en un entorno de producción.

2.1 Configuración despliegue frontend

2.1.1 Crear proyecto en Firebase Hosting





2.1.2 Configuración firebase primer despliegue

Instalación firebase

npm install -g firebase-tools

Iniciar sesión en el navegador

firebase login

Generar token para despliegue

firebase login:ci

Configuración de despliegue

firebase init hosting

- Seleccionar proyecto firebase (test en este caso)
- Seleccionar carpeta de versión de producción (build por defecto)
- Seleccionar Single Page App (SPA)

2.2 Crear archivo de despliegue

name: Nombre descriptivo del workflow.

name: Deploy to Firebase Hosting on merge

on.push: Dispara este workflow cada vez que haya un push (o merge) a la rama main.

on:

 push:

 branches:

 - main

Jobs: crea el job build_and_deploy.

runs-on: ubuntu-latest especifica que el job se ejecute en un runner de Ubuntu.

jobs:

 build_and_deploy:

 runs-on: ubuntu-latest

 steps:

Usa la acción oficial actions/checkout@v4 para clonar el repositorio en el runner y poder acceder al código durante la ejecución del runner.

 - name: Checkout repository

 uses: actions/checkout@v4

Instala Node.js (versión 22) en el runner y ajusta las variables de entorno (PATH, NODE_HOME).

```
- name: Setup Node.js
  uses: actions/setup-node@v3
  with:
    node-version: 22
```

Añade al final de .env una línea con tu variable VITE_APP_API_URL, cuyo valor se toma de los Secrets de Github.

```
- name: Crear archivo .env
  run: |
    echo "VITE_APP_API_URL=${{ secrets.VITE_APP_API_URL }}" >> .env
```

Ejecuta npm install para descargar todas las dependencias listadas en el package.json

```
- name: Install dependencies
  run: npm install
```

Lanza el script build que ejecuta vite build. Genera la carpeta build/ lista para producción.

```
- name: Build project
  run: npm run build
```

- uses: aprovecha la acción oficial de Firebase Extended para hosting.
- repoToken: se pasa el token para que la acción pueda interactuar con GitHub.
- firebaseServiceAccount: JSON de credenciales de la cuenta de servicio de Firebase (guardado como secreto FIREBASE_SERVICE_ACCOUNT_BBERRYCOMPANY).
- channelId: live: rama de despliegue en Firebase Hosting. live corresponde al release channel por defecto.
- projectId: bberrycompany: identifica el proyecto de Firebase al que se va a desplegar.

```
- name: Deploy to Firebase Hosting
```

```

uses: FirebaseExtended/action-hosting-deploy@v0
with:
  repoToken: {{ secrets.GITHUB_TOKEN }}
  firebaseServiceAccount: {{ secrets.FIREBASE_SERVICE_ACCOUNT_BBERRYCOMPANY }}
  channelId: live
  projectId: bberrycompany

```

3 Codificación Modelos Matemáticos

Para la codificación de los modelos matemáticos se debe contar con el dataset en formato cvs, como lenguaje de programación se empleó Python con ayuda de las siguientes librerías para cada uno de los modelos como se presenta en la siguiente tabla.

Programa	librerías
Modelo GAM	<ul style="list-style-type: none"> • pandas • numpy • pygam • scikit-learn • matplotlib • joblib
Modelo Gompertz	<ul style="list-style-type: none"> • numpy • matplotlib • scipy • pandas • pickle (incluida en la biblioteca estándar de Python)
Modelo logístico	<ul style="list-style-type: none"> • numpy • matplotlib • scipy • pandas
Correlación de kendall	<ul style="list-style-type: none"> • pandas • numpy • matplotlib • seaborn

En esta tabla se presentan los 4 programas y las librerías que se usaron en cada uno de ellos, además de poder realizar la instalación de todas las librerías con el siguiente comando de Python.

```

pip install pandas numpy pygam scikit-learn matplotlib joblib
scipy seaborn

```

Con esto ya instalado se puede proceder con la creación de los archivos .py en donde se describe el cuerpo de cada uno de ellos a continuación comenzando con el código completo del modelo GAM

3.1 ModeloGam.py

```
import pandas as pd
import numpy as np
from pygam import LinearGAM, s, f
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_squared_error
import matplotlib.pyplot as plt
import joblib

# Carga del dataset
df = pd.read_csv('dataset_con_poda.csv',
parse_dates=['timestamp'])

# Verificar carga correcta del dataset
print(df.head())

# Definimos X (predictores) y Y (variable objetivo)
X = df[['temp_high', 'temp_low', 'hum_rel', 'min_siembra',
'hum_sustr', 'lum', 'temp', 'poda_engrosamiento']]
y = df['altura']

# División en entrenamiento y prueba (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
# Definimos el modelo GAM
# Definimos el modelo GAM con splines cúbicos y un factor
categórico
gam = LinearGAM(
    s(0, n_splines=10) +
    s(1, n_splines=10) +
    s(2, n_splines=10) +
    s(3, n_splines=10) +
    s(4, n_splines=10) +
    s(5, n_splines=10) +
    s(6, n_splines=10) +
    f(7)
).fit(X_train, y_train)
# Guardar el modelo entrenado
joblib.dump(gam, 'gam_model.pkl')
# Cargar el modelo entrenado
gam = joblib.load('gam_model.pkl')
```

```

# Predicciones sobre conjunto de prueba
y_pred = gam.predict(X_test)

# Métricas de evaluación
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print(f"R2 (Porcentaje explicado por el modelo):
{r2*100:.2f}%")
print(f"RMSE (Error cuadrático medio): {rmse:.4f}")
#muestra de los terminos del modelo y el coeficiente
gam.summary()
print(gam.terms)

print(gam.coef_[-1])

fig, axs = plt.subplots(2, 4, figsize=(20, 10))

titles = ['temp_high', 'temp_low', 'hum_rel', 'min_siembra',
'hum_sustr', 'lum', 'temp', 'poda_engrosamiento']

for i, ax in enumerate(axs.flat):
    XX = gam.generate_X_grid(term=i)
    ax.plot(XX[:, i], gam.partial_dependence(term=i, X=XX))
    ax.plot(XX[:, i], gam.partial_dependence(term=i, X=XX,
width=0.95)[1], c='r', ls='--')
    ax.set_title(f"Efecto parcial de {titles[i]}")
    ax.grid(True)
    # Ocultar último subplot (no utilizado)
plt.tight_layout()
plt.show()

partial_data = {}

# Iterar sobre los predictores
for i, feature in enumerate(['temp_high', 'temp_low',
'hum_rel', 'min_siembra', 'hum_sustr', 'lum', 'temp',
'poda_engrosamiento']):
    XX = gam.generate_X_grid(term=i, n=100) # Generar puntos
x
    pdep, confi = gam.partial_dependence(term=i, X=XX,
width=0.95)
    # Ajustar dimensiones con la transposición
    partial_data[feature] = pd.DataFrame({
        feature: XX[:, i],

```

```

        'partial_effect': pdep,
        'confidence_lower': confi[:, 0], # Intervalo inferior
        'confidence_upper': confi[:, 1] # Intervalo superior
    })
# Guardar en CSV
for feature, df in partial_data.items():
    df.to_csv(f'efectos/partial_effect_{feature}.csv',
index=False)

```

3.1.1 Explicación por Fracciones del Código

- **Importación de librerías**

```

import pandas as pd
import numpy as np
from pygam import LinearGAM, s, f
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_squared_error
import matplotlib.pyplot as plt
import joblib

```

- **Carga del dataset**

```

df = pd.read_csv('nombre_del_dataset.csv',
parse_dates=['timestamp'])
print(df.head())

```

- **Definición de variables predictoras y variable objetivo**

```

X = df[['temp_high', 'temp_low', 'hum_rel', 'min_siembra',
'hum_sustr', 'lum', 'temp', 'poda_engrosamiento']]
y = df['altura']

```

Donde:

X: contiene 7 variables numéricas y una categórica (poda_engrosamiento).

y: es la variable objetivo a predecir: **la altura**.

- **División del dataset**

```

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

```

Se divide el dataset en:

- 80% para entrenamiento (**X_train, y_train**)

- 20% para prueba (**X_test, y_test**)
- **random_state** asegura reproducibilidad.

- **Creación y entrenamiento del modelo GAM**

```
gam = LinearGAM(
    s(0, n_splines=10) +
    s(1, n_splines=10) +
    s(2, n_splines=10) +
    s(3, n_splines=10) +
    s(4, n_splines=10) +
    s(5, n_splines=10) +
    s(6, n_splines=10) +
    f(7)
).fit(X_train, y_train)
```

Se crea un modelo GAM lineal con:

- Términos suavizados (**s()**) para las primeras 7 variables numéricas.
- Un término categórico (**f(7)**) para poda_engrosamiento.
- Cada **s()** usa 10 splines.
- **.fit()** entrena el modelo.

- **Guarda y carga el modelo en un archivo .pkl**

```
joblib.dump(gam, 'gam_model.pkl')
gam = joblib.load('gam_model.pkl')
```

- **Realizar predicciones y evaluar la calidad del modelo**

```
y_pred = gam.predict(X_test)
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
```

Se hacen predicciones sobre el conjunto de prueba.

Se evalúa el modelo usando:

- **R²**: proporción de varianza explicada por el modelo.
- **RMSE**: error cuadrático medio.

- **Mostrar resumen del modelo**

```
gam.summary()
print(gam.terms)
print(gam.coef_[-1])
```

- **.summary()**: muestra estadísticos del modelo por cada término.
- **.terms**: muestra los términos usados (splines, factores...).

- **.coef_-1]:** imprime el último coeficiente (probablemente un efecto de la categoría poda_engrosamiento).

- **Visualizar efector parciales**

```
fig, axs = plt.subplots(2, 4, figsize=(20, 10))
titles = ['temp_high', 'temp_low', 'hum_rel', 'min_siembra',
'hum_sustr', 'lum', 'temp', 'poda_engrosamiento']

for i, ax in enumerate(axs.flat):
    XX = gam.generate_X_grid(term=i)
    ax.plot(XX[:, i], gam.partial_dependence(term=i, X=XX))
    ax.plot(XX[:, i], gam.partial_dependence(term=i, X=XX,
width=0.95)[1], c='r', ls='--')
    ax.set_title(f"Efecto parcial de {titles[i]}")
    ax.grid(True)

plt.tight_layout()
plt.show()
```

Se grafican los efectos parciales para cada variable, de manera que **.generate_X_grid()** crea valores posibles para evaluar la dependencia parcial.

Se grafican:

- **Línea del efecto parcial.**
- **Banda de confianza al 95%.**

- **Guardar efectos parciales en archivos CVS**

```
partial_data = {}

for i, feature in enumerate(['temp_high', 'temp_low',
'hum_rel', 'min_siembra', 'hum_sustr', 'lum', 'temp',
'poda_engrosamiento']):
    XX = gam.generate_X_grid(term=i, n=100)
    pdep, confi = gam.partial_dependence(term=i, X=XX,
width=0.95)
    partial_data[feature] = pd.DataFrame({
        feature: XX[:, i],
        'partial_effect': pdep,
        'confidence_lower': confi[:, 0],
        'confidence_upper': confi[:, 1]
    })
```

}}

```
PS C:\Users\jhona\Desktop\Modelos> & C:/Users/jhona/AppData/Local/Microsoft/WindowsApps/python3.12.exe c:/Users/jhona/Desktop/Modelos/GAM/gam.py
timestamp temp_high temp_low hum_rel min_siembra hum_sustr lum temp altura poda_engrosamiento
0 2024-01-01 00:00:00 22.0 12.0 91.029 87540 100.0 257.563 17.681 14.333 0
1 2024-01-01 00:10:00 22.0 12.0 91.311 87550 100.0 308.760 17.648 14.333 0
2 2024-01-01 00:20:00 22.0 12.0 91.576 87560 100.0 322.240 17.615 14.333 0
3 2024-01-01 00:30:00 22.0 12.0 91.832 87570 100.0 329.950 17.581 14.333 0
4 2024-01-01 00:40:00 22.0 12.0 91.790 87580 100.0 325.857 17.548 14.333 0
R² (Porcentaje explicado por el modelo): 97.83%
RMSE (Error cuadrático medio): 1.8589
LinearGAM
=====
Distribution: NormalDist Effective DoF: 56.7993
Link Function: IdentityLink Log Likelihood: -118527.1906
Number of Samples: 50934 AIC: 237169.9798
AICc: 237170.1134
GCV: 3.5502
Scale: 3.5521
Pseudo R-Squared: 0.9778
=====
Feature Function Lambda Rank EDoF P > x Sig. Code
=====
s(0) [0.6] 10 8.8 1.11e-16 ***
s(1) [0.6] 10 8.5 1.11e-16 ***
s(2) [0.6] 10 7.6 1.11e-16 ***
s(3) [0.6] 10 8.5 1.11e-16 ***
s(4) [0.6] 10 7.8 6.66e-16 ***
s(5) [0.6] 10 7.6 7.35e-01
s(6) [0.6] 10 7.0 3.59e-04 ***
f(7) [0.6] 2 1.0 1.11e-16 ***
intercept 1 0.0 1.11e-16 ***
=====
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

WARNING: Fitting splines and a linear function to a feature introduces a model identifiability problem
which can cause p-values to appear significant when they are not.

WARNING: p-values calculated in this manner behave correctly for un-penalized models or models with
known smoothing parameters, but when smoothing parameters have been estimated, the p-values
are typically lower than they should be, meaning that the tests reject the null too readily.
c:\Users\jhona\Desktop\Modelos\GAM\gam.py:48: UserWarning: KNOWN BUG: p-values computed in this summary are likely much smaller than they should be.

Please do not make inferences based on these values!

Collaborate on a solution, and stay up to date at:
github.com/dswah/pyGAM/issues/163

gam.summary()
s(0) + s(1) + s(2) + s(3) + s(4) + s(5) + s(6) + f(7) + intercept
18.233847794039207
```

```
for feature, df in partial_data.items():
    df.to_csv(f'efectos/partial_effect_{feature}.csv',
index=False)
```

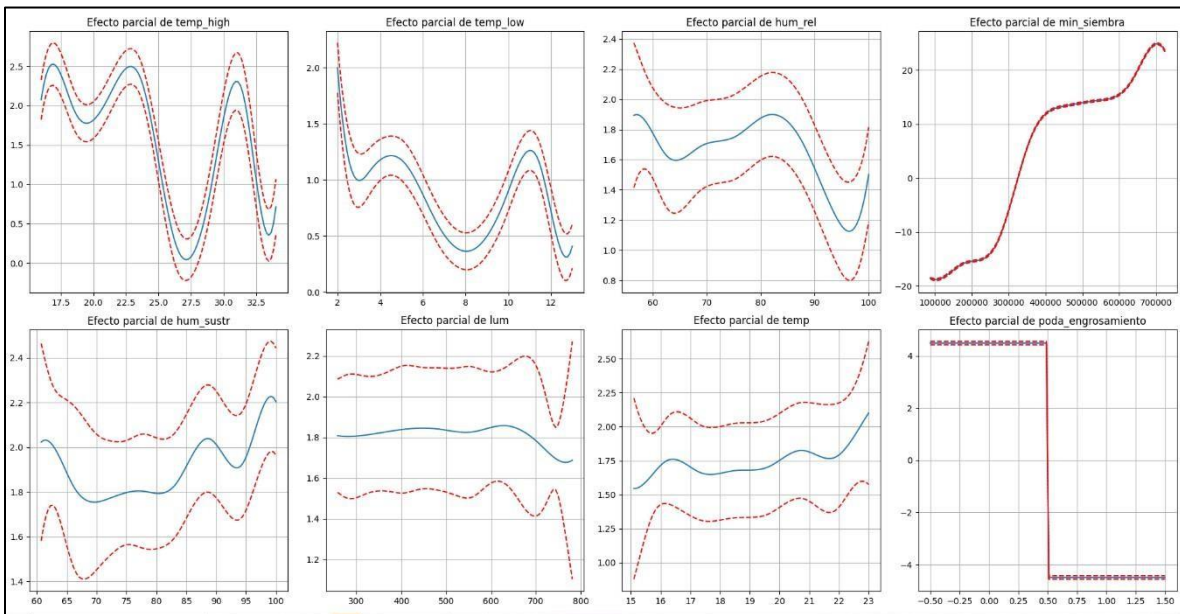
Se genera una **tabla** con:

- Valores del predictor.
- Efecto parcial.
- Intervalo de confianza.

Luego se guarda cada tabla en un archivo **CSV** por separado en la carpeta efectos.

3.1.2 Resultados Obtenidos

- Consola
- Gráficas



3.2 ModeloGompertz.py

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
import pandas as pd
import pickle # Para guardar el modelo

# Cargar el dataset completo
df = pd.read_csv('dataset_con_poda.csv',
parse_dates=['timestamp'])

# Agrupar por día y calcular el promedio de 'altura' por día
df['fecha'] = df['timestamp'].dt.date # Extraer solo la fecha
df_por_dia =
df.groupby('fecha')['altura'].mean().reset_index()

# Crear x_data (índice de días) y y_data (altura promedio por
día)
```

```

x_data = np.arange(len(df_por_dia)) # 0, 1, 2, ...
y_data = df_por_dia['altura'].to_numpy()

# Función de Gompertz con offset
def gompertz(x, L, k, x0, y0):
    return y0 + L * np.exp(-np.exp(-k * (x - x0)))

# Estimar parámetros iniciales
y_max = y_data.max()
y_min = y_data.min()
p0 = [y_max - y_min, 0.1, x_data.mean(), y_min]

try:
    # Ajuste del modelo
    popt, _ = curve_fit(gompertz, x_data, y_data, p0=p0,
maxfev=10000)
    L, k, x0, y0 = popt

    print("Parámetros ajustados:")
    print(f"L (altura adicional): {L:.3f}")
    print(f"k (tasa crecimiento): {k:.3f}")
    print(f"x0 (inflexión): {x0:.3f}")
    print(f"y0 (offset base): {y0:.3f}")

    # Guardar el modelo en un archivo .pkl
    with open('modelo_gompertz.pkl', 'wb') as f:
        pickle.dump({'L': L, 'k': k, 'x0': x0, 'y0': y0}, f)
    print("Modelo guardado como 'modelo_gompertz.pkl'")

    # Predicción
    y_pred = gompertz(x_data, L, k, x0, y0)

    # Error cuadrático medio
    mse = np.mean((y_data - y_pred) ** 2)
    print(f"Error cuadrático medio: {mse:.6f}")

    # Visualización
    plt.figure(figsize=(10, 6))
    plt.scatter(x_data, y_data, label='Datos (promedio por
día)', color='blue', alpha=0.5)
    plt.plot(x_data, y_pred, label='Ajuste Gompertz con
offset', color='red')
    plt.xlabel('Días desde el inicio')
    plt.ylabel('Altura')
    plt.title('Ajuste de la función Gompertz a datos de altura
(sin filtro de fecha)')

```

```

plt.legend()
plt.grid(True)
plt.show()

#Excepción para errores de ajuste
except RuntimeError:
    print("El ajuste falló. Revisa los datos o ajusta los
valores iniciales (p0).")

```

3.2.1 Explicación por Fracciones de Código

- **Importación de librerías**

```

import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
import pandas as pd
import pickle

```

- **numpy:** operaciones numéricas.
- **matplotlib.pyplot:** graficación.
- **curve_fit:** ajuste de curvas no lineales.
- **pandas:** manejo de datos.
- **pickle:** serialización (guardar objetos Python, como el modelo ajustado).

- **Carga del dataset**

```

df = pd.read_csv('nombre_Data_Set.csv',
parse_dates=['timestamp'])

```

- **Agrupación de los datos por día**

```

x_data = np.arange(len(df_por_dia))
y_data = df_por_dia['altura'].to_numpy()

```

x_data: es una secuencia de días, como **[0, 1, 2, ...]**.

y_data: alturas promedio por día, convertidas a array **NumPy**.

- **Definición de gompertz**

```
def gompertz(x, L, k, x0, y0):
    return y0 + L * np.exp(-np.exp(-k * (x - x0)))
```

Función de crecimiento sigmoideal asimétrica.

Parámetros:

- **L**: altura adicional (máxima esperada menos el offset).
- **k**: tasa de crecimiento.
- **x0**: punto de inflexión (día donde el crecimiento cambia de acelerado a lento).
- **y0**: altura inicial (offset base).
- **Estimación de parámetros iniciales**

```
y_max = y_data.max()
y_min = y_data.min()
p0 = [y_max - y_min, 0.1, x_data.mean(), y_min]
```

Propuesta inicial para **curve_fit**:

- **L** \approx $y_{\max} - y_{\min}$
- **k** = 0.1: valor estándar como tasa de crecimiento.
- **x0** = promedio del eje x: día medio.
- **y0** = altura mínima: offset.

- Ajuste del modelo y desempaquetamiento de los parámetros

```
popt, _ = curve_fit(gompertz, x_data, y_data, p0=p0,
maxfev=10000)
```

Ajusta los parámetros reales (**popt**) que mejor se ajustan a los datos.
maxfev=10000: número máximo de evaluaciones de función para asegurar convergencia.

- **Mostrar parámetros y guardar el modelo**

```
print("Parámetros ajustados: ...")
with open('modelo_gompertz.pkl', 'wb') as f:
    pickle.dump({'L': L, 'k': k, 'x0': x0, 'y0': y0}, f)
```

- **Evaluación del modelo**

```
y_pred = gompertz(x_data, L, k, x0, y0)
mse = np.mean((y_data - y_pred) ** 2)
```

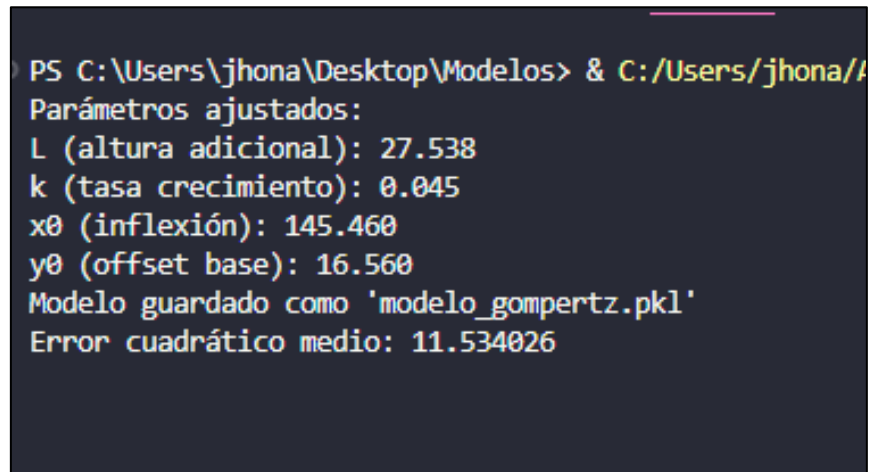
Se predicen las alturas usando la función con los parámetros ajustados. Se calcula el **Error Cuadrático Medio (MSE)** entre los datos reales y las predicciones.

- **visualización**

```
plt.figure(figsize=(10, 6))
plt.scatter(x_data, y_data, label='Datos', color='blue',
            alpha=0.5)
plt.plot(x_data, y_pred, label='Gompertz', color='red')
plt.xlabel('Días desde el inicio')
plt.ylabel('Altura')
plt.title('Ajuste de la función Gompertz')
plt.legend()
plt.grid(True)
plt.show()
```

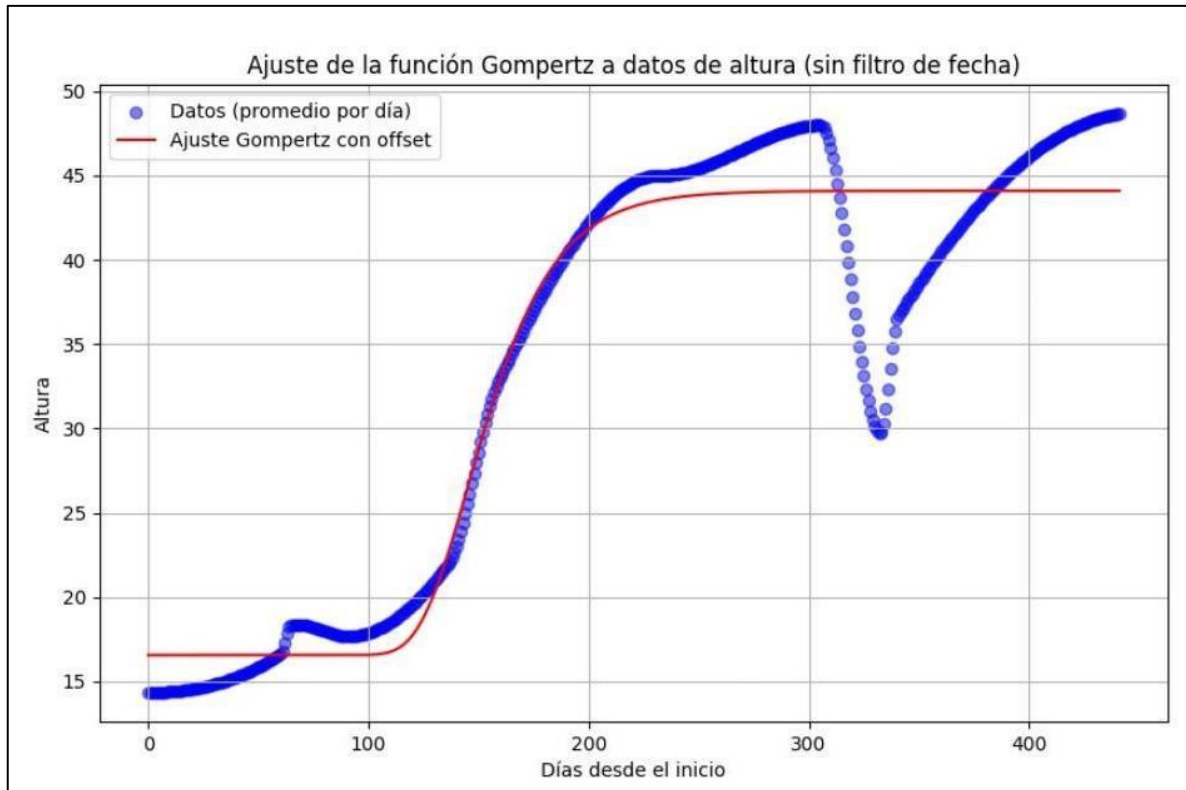
3.2.2 Resultados

- **Consola**



```
PS C:\Users\jhona\Desktop\Modelos> & C:/Users/jhona/A
Parámetros ajustados:
L (altura adicional): 27.538
k (tasa crecimiento): 0.045
x0 (inflexión): 145.460
y0 (offset base): 16.560
Modelo guardado como 'modelo_gompertz.pkl'
Error cuadrático medio: 11.534026
```

- Gráfico



3.3 ModeloLogistico.py

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
import pandas as pd

# Cargar el dataset
df = pd.read_csv('dataset_con_poda.csv',
parse_dates=['timestamp'])

# Filtrar solo los datos donde 'poda_engrosamiento' sea igual
a 1
df_filtrado = df[df['poda_engrosamiento'] == 1].copy()

# Verificar si hay datos después del filtro
if not df_filtrado.empty:
    # Agrupar por día y calcular el promedio de 'altura' por
día
    df_filtrado['fecha'] = df_filtrado['timestamp'].dt.date
```

```

df_por_dia =
df_filtrado.groupby('fecha')['altura'].mean().reset_index()

# Crear x_data (índice de días) y y_data (altura promedio
por día)
x_data = np.arange(len(df_por_dia))
y_data = df_por_dia['altura'].to_numpy()

# Definir la función logística con offset
def logistica(x, L, k, x0, y0):
    return y0 + L / (1 + np.exp(-k * (x - x0)))

# Estimar parámetros iniciales
y_max = y_data.max()
y_min = y_data.min()
p0 = [y_max - y_min, 0.1, x_data.mean(), y_min]

try:
    # Ajustar la función logística
    popt, _ = curve_fit(logistica, x_data, y_data, p0=p0,
maxfev=10000)
    L, k, x0, y0 = popt

    print("\nParámetros ajustados:")
    print(f"Altura adicional (L): {L:.3f}")
    print(f"Tasa de crecimiento (k): {k:.3f}")
    print(f"Parámetro de inflexión (x0): {x0:.3f}")
    print(f"Valor base (y0): {y0:.3f}")

    # Imprimir fórmula ajustada
    print("\n Función ajustada (logística):")
    print(f"y(x) = {y0:.3f} + {L:.3f} / (1 + exp(-{k:.3f}
* (x - {x0:.3f})))")

    # Predicciones
    y_pred = logistica(x_data, L, k, x0, y0)

    # Error cuadrático medio
    mse = np.mean((y_data - y_pred) ** 2)
    print(f"\nError cuadrático medio (MSE): {mse:.6f}")

    # Guardar parámetros
    df_parametros = pd.DataFrame({
        'Parametro': ['L', 'k', 'x0', 'y0', 'MSE'],
        'Valor': [L, k, x0, y0, mse]
    })

```

```

df_parametros.to_csv('parametros_logistica.csv',
index=False)

# Guardar predicciones
df_predicciones = pd.DataFrame({
    'Día': x_data,
    'Fecha': df_por_dia['fecha'],
    'Altura Real': y_data,
    'Altura Predicha': y_pred
})
df_predicciones.to_csv('predicciones_logistica.csv',
index=False) # Guardar predicciones

# Visualización
plt.figure(figsize=(12, 7))
plt.scatter(x_data, y_data, label='Datos reales
(promedio diario)',
            color='green', s=60, edgecolors='black',
alpha=0.8)
plt.plot(x_data, y_pred, label='Modelo logístico',
color='purple', linewidth=2)
plt.scatter(x_data, y_pred, label='Puntos del modelo',
color='orange', s=40, marker='x')
plt.xlabel('Días desde el inicio')
plt.ylabel('Altura')
plt.title('Ajuste de la función logística a datos de
altura con offset')
plt.grid(True)
plt.legend(loc='upper left')

# Mostrar parámetros en la gráfica
texto_parametros = (
    f"Parámetros ajustados:\n"
    f"L: {L:.3f}\n"
    f"k: {k:.3f}\n"
    f"x0: {x0:.3f} días\n"
    f"y0: {y0:.3f}\n"
    f"ECM: {mse:.6f}"
)
plt.text(0.98, 0.02, texto_parametros,
transform=plt.gca().transAxes,
fontSize=10,
verticalalignment='bottom',
horizontalalignment='right',
bbox=dict(facecolor='white', alpha=0.85,
boxstyle='round'))

```

```

plt.tight_layout()
plt.show()
#Manejo de errores
except RuntimeError:
    print(" El ajuste falló. Revisa los datos o ajusta los
valores iniciales (p0).")
else:
    print(" No hay datos con poda_engrosamiento = 1 para
realizar el ajuste.")

```

3.3.1 Explicación por Fracciones de Código

- **Importación de librerías**

```

import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
import pandas as pd

```

- **numpy:** para manejo de arrays y cálculos numéricos.
- **matplotlib.pyplot:** para graficar.
- **scipy.optimize.curve_fit:** para ajustar funciones a datos.
- **pandas:** para manejo de datos tabulares.

- **Carga del dataset, filtrado de datos, agrupación por día**

```

df = pd.read_csv('dataset_con_poda.csv',
parse_dates=['timestamp'])
df_filtrado = df[df['poda_engrosamiento'] == 0].copy()
df_filtrado['fecha'] = df_filtrado['timestamp'].dt.date
df_por_dia =
df_filtrado.groupby('fecha')['altura'].mean().reset_index()

```

- **Preparación de los datos para realizar ajuste**

```

x_data = np.arange(len(df_por_dia))
y_data = df_por_dia['altura'].to_numpy()

```

x_data: vector de enteros representando los días (0, 1, 2...).

y_data: vector de alturas promedio por día.

- **Definición de la función logística**

```

def logistica(x, L, k, x0, y0):

```

```
return y0 + L / (1 + np.exp(-k * (x - x0)))
```

Esta es una función logística comúnmente usada para modelar crecimiento:

- **L**: incremento máximo en altura.
- **k**: tasa de crecimiento.
- **x0**: día de inflexión (cuando más rápido crece).
- **y0**: Altura inicial.

- **Estimacion de parametros**

```
y_max = y_data.max()
y_min = y_data.min()
p0 = [y_max - y_min, 0.1, x_data.mean(), y_min]
```

Se calculan valores iniciales para el ajuste:

- **L** \approx **y_max - y_min**: cambio total de altura.
- **k = 0.1**: **suposición inicial de la pendiente**.
- **x0**: punto medio de los datos.
- **y0**: altura inicial mínima.

- **Ajuste de la curva y resultado del modelo**

```
popt, _ = curve_fit(logistica, x_data, y_data, p0=p0,
maxfev=10000)
L, k, x0, y0 = popt
```

Ajusta la función logística a los datos (**x_data**, **y_data**) usando los parámetros iniciales **p0**.

maxfev=10000: aumenta el número de iteraciones permitidas.

- **Predicción y error**

```
y_pred = logistica(x_data, L, k, x0, y0)
mse = np.mean((y_data - y_pred) ** 2)
```

Se calcula la curva predicha con los parámetros ajustados.

Se mide el **Error Cuadrático Medio (MSE)** entre las alturas reales y las predichas.

- **Guardado de resultados en archivos csv**

```
df_parametros.to_csv('parametros_logistica.csv', index=False)
df_predicciones.to_csv('predicciones_logistica.csv',
index=False)
```

- **visualización**

```
plt.figure(figsize=(12, 7))
    plt.scatter(x_data, y_data, label='Datos reales
(promedio diario)',
                color='green', s=60, edgecolors='black',
alpha=0.8)
    plt.plot(x_data, y_pred, label='Modelo logístico',
color='purple', linewidth=2)
    plt.scatter(x_data, y_pred, label='Puntos del modelo',
color='orange', s=40, marker='x')
    plt.xlabel('Días desde el inicio')
    plt.ylabel('Altura')
    plt.title('Ajuste de la función logística a datos de
altura con offset')
    plt.grid(True)
    plt.legend(loc='upper left')

# Mostrar parámetros en la gráfica
texto_parametros = (
    f"Parámetros ajustados:\n"
    f"L: {L:.3f}\n"
    f"k: {k:.3f}\n"
    f"x0: {x0:.3f} días\n"
    f"y0: {y0:.3f}\n"
    f"ECM: {mse:.6f}"
)
plt.text(0.98, 0.02, texto_parametros,
        transform=plt.gca().transAxes,
        fontsize=10,
        verticalalignment='bottom',
        horizontalalignment='right',
        bbox=dict(facecolor='white', alpha=0.85,
boxstyle='round'))

plt.tight_layout()
plt.show()
```

- **Manejo de errores y excepciones**

```
#Manejo de errores
    except RuntimeError:
        print(" El ajuste falló. Revisa los datos o ajusta los
valores iniciales (p0).")
    else:
```

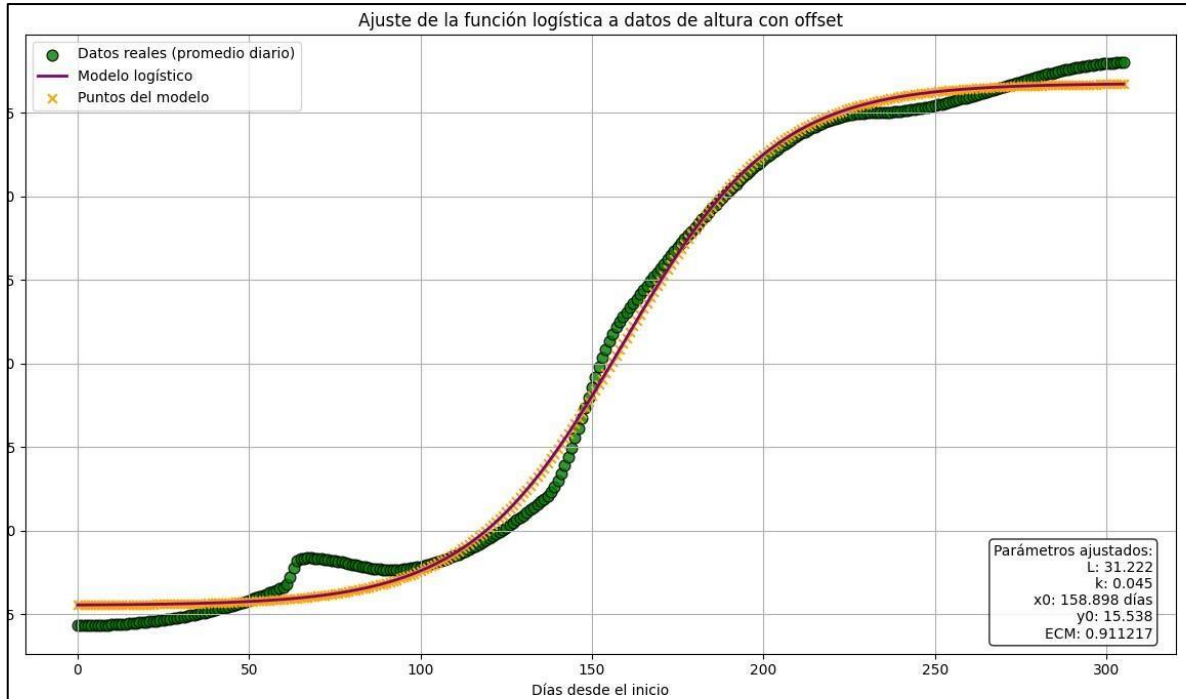
```
print(" No hay datos con poda_engrosamiento = 1 para  
realizar el ajuste.")
```

3.3.2 Resultados

- Consola

```
Parámetros ajustados:  
Altura adicional (L): 31.222  
Tasa de crecimiento (k): 0.045  
Parámetro de inflexión (x0): 158.898  
Valor base (y0): 15.538  
  
Función ajustada (logística):  
 $y(x) = 15.538 + 31.222 / (1 + \exp(-0.045 * (x - 158.898)))$   
  
Error cuadrático medio (MSE): 0.911217
```

- Gráfica



3.4 CorrelacionKendall.py

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('dataset_con_poda.csv',
parse_dates=['timestamp'])
# Verificar carga correcta
print(df.head())
df_numeric = df.drop('timestamp', axis=1)
# Calcular las matrices de correlación
corr_kendall = df_numeric.corr(method='kendall')
# Gráfico 3: Mapa de calor para Kendall
plt.figure(figsize=(8, 6))
sns.heatmap(corr_kendall, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlación de Kendall')
plt.tight_layout()
plt.show()
# Imprimir las matrices para referencia
print("\nMatriz de correlación (Kendall):")
print(corr_kendall)
```

3.4.1 Explicación por Fracciones de Código

- Importación de librerías

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

pandas (pd) se usa para manipular datos tabulares (como archivos .csv).

numpy (np) se usa para operaciones numéricas, aunque en este código en particular no se utiliza directamente.

matplotlib.pyplot (plt) y **seaborn (sns)** son librerías de visualización:

matplotlib permite crear gráficos personalizables.

seaborn es una capa sobre **matplotlib** que simplifica la creación de gráficos estadísticos atractivos, como **mapas de calor**.

- Carga del dataset a evaluar

```
df = pd.read_csv('dataset_con_poda.csv',
parse_dates=['timestamp'])
```

- **Verificación de datos y limpieza de columnas**

```
print(df.head())
df_numeric = df.drop('timestamp', axis=1)
```

- **Calculo correlacion de kendall**

```
corr_kendall = df_numeric.corr(method='kendall')
```

Calcula la matriz de correlación usando el método de **Kendall (τ)**.

Kendall es un coeficiente no paramétrico que mide dependencia ordinal entre variables.

- **Valores cercanos a 1:** correlación positiva fuerte.
- **Valores cercanos a -1:** correlación negativa fuerte.
- **Valores cerca de 0:** sin correlación.

- **Grafico del mapa de calor**

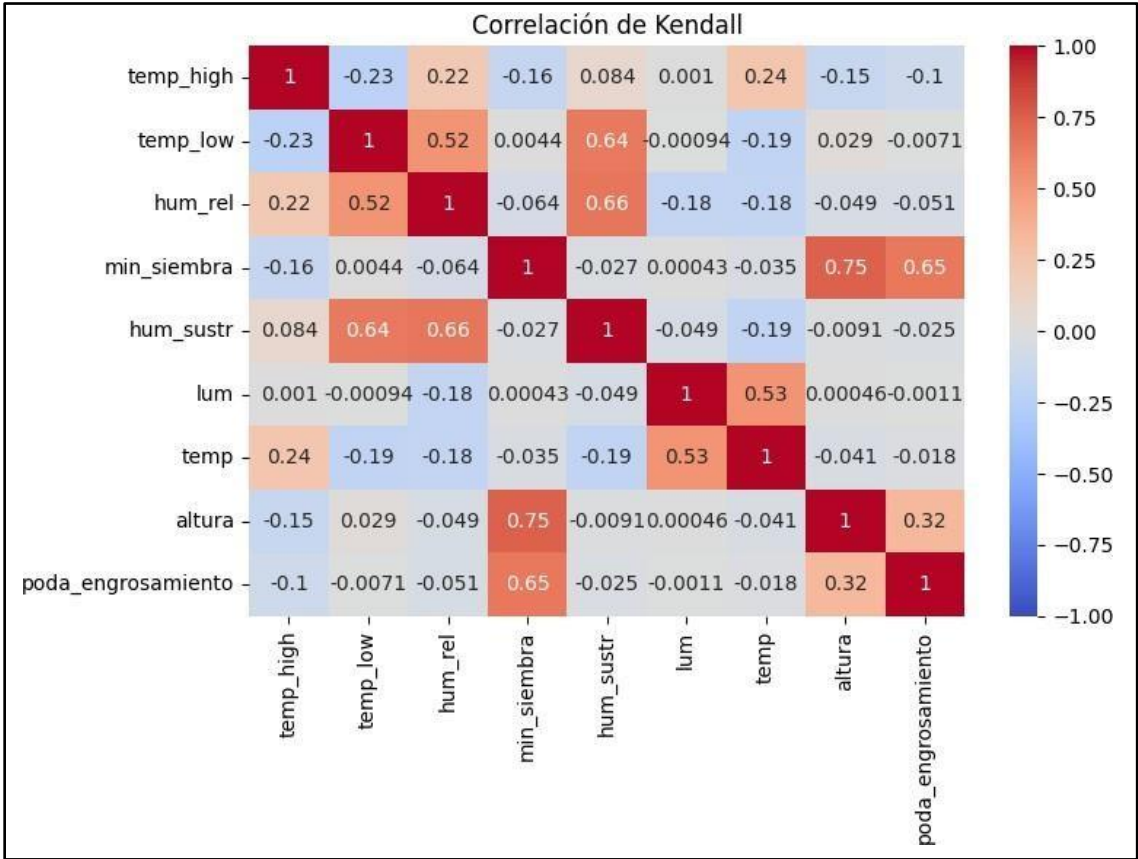
```
plt.figure(figsize=(8, 6))
sns.heatmap(corr_kendall, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlación de Kendall')
plt.tight_layout()
plt.show()
```

3.4.2 Resultados

- **Consola**

```
timestamp temp_high temp_low hum_rel min_siembra hum_sustr lum temp altura poda_engrosamiento
0 2024-01-01 00:00:00 22.0 12.0 91.029 87540 100.0 257.563 17.681 14.333 0
1 2024-01-01 00:10:00 22.0 12.0 91.311 87550 100.0 308.760 17.648 14.333 0
2 2024-01-01 00:20:00 22.0 12.0 91.576 87560 100.0 322.240 17.615 14.333 0
3 2024-01-01 00:30:00 22.0 12.0 91.832 87570 100.0 329.950 17.581 14.333 0
4 2024-01-01 00:40:00 22.0 12.0 91.790 87580 100.0 325.857 17.548 14.333 0
```

- **Gráfica**



DATOS DE CONTACTO SOPORTE APLICACIÓN

Extensiones telefónicas: No Aplica
correo: company.bberry@gmail.com

DIRECCIÓN DE SISTEMAS Y TECNOLOGÍA

Ariel Camilo Sánchez López
Cargo: Pregrado Ingeniería de Sistemas y Computación

Jhonatan Murcia Posso
Cargo: Pregrado Ingeniería de Sistemas y Computación

ÁREA DE DESARROLLO DE SISTEMAS DE INFORMACIÓN



Título

MODELOS Y TÉCNICAS MATEMÁTICAS QUE DESCRIBEN EL CRECIMIENTO DE PLANTAS DE ARÁNDANOS BILOXY MEDIANTE COMPARACIÓN DE VARIABLES AMBIENTALES.

Nombre Autor 1: Oscar Javier Morera Zárate

Grupo de investigación GISTFA Universidad de Cundinamarca, Facatativa Cund, Colombia.

ORCID: <https://orcid.org/0000-0001-6160-5124>

Correo: institucional: Omorera@ucundinamarca.edu.co

Perfil: Docente Investigador Principal

Nombre Autor 2: Gina Maribel Valenzuela Sabogal

Grupo de investigación GISTFA Universidad de Cundinamarca, Facatativa Cund, Colombia.

ORCID: <https://orcid.org/0000-0002-2833-1579>

Correo: gvalenzuela@ucundinamarca.edu.co

Perfil: Docente Coinvestigador

Nombre Autor 3: Ariel Camilo Sánchez López

Grupo de investigación GISTFA Universidad de Cundinamarca, Facatativa Cund, Colombia.

ORCID: <https://orcid.org/0009-0003-0794-8187>

Correo: arielcsanchez@ucundinamarca.edu.co

Perfil: Estudiante investigador

Nombre Autor 4: Jhonatan Murcia Posso

Grupo de investigación GISTFA Universidad de Cundinamarca, Facatativa Cund, Colombia.

ORCID: <https://orcid.org/0009-0009-8645-2596>

Correo: jmurciap@ucundinamarca.edu.co

Perfil: Estudiante investigador

Resumen

El estudio que se presenta se enfoca en el desarrollo de un modelo matemático para describir el crecimiento de arándanos Biloxi, mediante la comparación de variables ambientales en sistemas cerrados y abiertos. El objetivo principal es investigar, desarrollar e implementar un modelo matemático que por medio de herramientas tecnológicas describa la relación entre el crecimiento de la planta de arándano Biloxi y las variables ambientales en los diferentes sistemas de cultivo.

El contexto del estudio abarca una revisión exhaustiva de la literatura a nivel global, regional y nacional, con el fin de establecer un conocimiento sólido sobre los aspectos relevantes para el desarrollo del proyecto. Se exploran diversas técnicas y metodologías de modelado, tanto estadísticas como mecanicistas, con el propósito de determinar cuál se adapta mejor a los datos obtenidos sobre el ambiente y el crecimiento de la planta.

Se destaca la importancia de considerar modelos no lineales, como los modelos aditivos generalizados (GAM) y los modelos basados en funciones sigmoides, para capturar adecuadamente los patrones de crecimiento del arándano Biloxi. Además, se discuten técnicas estadísticas multivariantes, como el análisis de varianza multivariante (MANOVA), el análisis discriminante y el análisis de correlación canónica, para comparar y describir las relaciones entre las variables ambientales y el crecimiento de la planta.

El estudio sugiere el desarrollo de un modelo preciso y adaptado a las necesidades del cultivo puede mejorar la comprensión de los factores o variables relevantes que influyen en el crecimiento y el comportamiento de las plantas en diferentes ambientes.

Palabras clave: Modelos Mecanicistas, Modelos Comparativos y descriptivos, Técnicas estadísticas, Ambientes Controlados, Variables de crecimiento.

Abstract

Es la traducción del resumen. Para esto, la Universidad Incca de Colombia se encargará de hacer la traducción.

Keywords: Es la traducción de las palabras clave al inglés. Para esto, la Universidad Incca de Colombia se encargará de hacer la traducción.

Introducción

El propósito de realizar un modelo matemático que permita describir el crecimiento de una planta de arándanos de variedad biloxy por medio de variables ambientales es una solución que permite dar un paso más hacia una agricultura de precisión que permita al agricultor tener una idea más clara de cómo los diferentes factores ambientales afectan al crecimiento de las plantas y el aprovechamiento de los recursos como fertilizantes, agua y entre otros reduciendo el desperdicio de estos, esto se comprueba por medio de diferentes ambientes en los que se estudia y compara el comportamiento de las plantas y los recursos empleados para el correcto crecimiento de las mismas.

En el desarrollo de la investigación se encuentran diferentes modelos matemáticos que brindan un enfoque de desarrollo y aplicación más acertado a lo que se busca entre estos se destacan los modelos mecanicistas, los modelos mecanicistas son herramientas conceptuales que descomponen sistemas complejos en partes simples y describen sus interacciones utilizando leyes físicas y matemáticas (Machamer, P., Darden, L., & Craver, C. F. 2000), por otro lado también se plantean los modelos logistas basado en lo que propone (Hilty, 2021) los modelos logísticos pueden ser más apropiados para describir el crecimiento. Por último se busca una descripción y comparación que permita identificar cuáles son los factores que afectan más al crecimiento con base al ambiente en donde se encuentra la planta, para esto se indaga sobre modelos descriptivos lo cuales según Leandro et al. (2019), "los modelos descriptivos permiten explorar y comprender la estructura subyacente de los datos, identificar patrones y relaciones entre variables, y proporcionar una representación visual efectiva de la información" (p. 23), por otro lado los modelos comparativos que permiten tener una mejor veracidad de los datos obtenidos como señala Manzano et al. (2019), "los modelos comparativos buscan determinar si las diferencias observadas entre grupos son estadísticamente significativas o si, por el contrario,

podrían deberse al azar" (p. 49). Destacando técnicas como ANOVA para un grupo de diferentes que es una técnica de análisis multivariable y entre otras que se describen en el artículo.

Finalizando se busca tener una base sólida con cada uno de los modelos y técnicas investigadas hasta el momento dando como resultado las variables relevantes y el comportamiento de estas en los diferentes ambientes con el crecimiento, apuntando a una agricultura de precisión que ayude al sector productor de arándanos biloxy a un mejor control y distribución de recursos dando como resultado una producción más estable.

Método

La investigación que se está realizando se presenta como un caso de estudio sobre el comportamiento y el crecimiento de arándanos biloxy en ambientes cerrados y abiertos de recirculación de agua dando paso a una investigación experimental de diferentes ambientes que permiten tener una relación de las diferentes variables y como estas afectan a la planta.

El diseño que se busca implementar es un modelo comparativo y descriptivo, ya que por un lado se comparan las variables ambientales y el crecimiento de las plantas en los diferentes ambientes, lo que implica también un enfoque comparativo buscando describir y modelar la relación entre estas variables. La población de estudio son 4-5 plantas de arándanos de variedad biloxy que se encuentran en espacios cerrados (invernadero y ambiente controlado) y abierto (ambiente normal) donde se recolectan datos de diferentes variables en tiempo real como de temperatura, humedad, luz y entre otras que son importantes para un buen desarrollo de la planta, para poder medir estos datos se implementan sensores y actuadores que por medio de técnicas de análisis permitan dar datos más limpios que serán entregados al modelo para ser procesados

Por último, se describen las fases de estudio que son las siguientes

1. **Revisión de literatura y planificación:** Realizar una revisión exhaustiva de la literatura relevante y planificar el diseño experimental, la recolección de datos y los análisis a realizar.
2. **Preparación del entorno de cultivo:** Establecer los sistemas cerrados (invernadero y ambiente controlado) y abiertos (ambiente normal) para el cultivo de las plantas de arándano Biloxi.

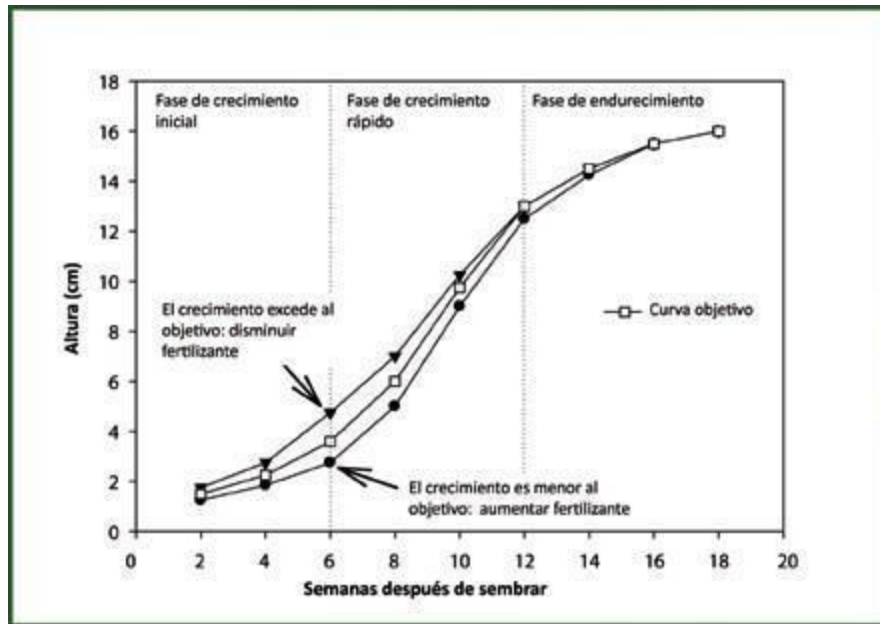
3. **Recolección de datos:** Monitorear y recolectar datos sobre las variables ambientales y el crecimiento de las plantas en los sistemas de cultivo durante un período de tiempo completo.
4. **Análisis de datos:** Aplicar técnicas estadísticas y de modelado matemático para analizar los datos recolectados y desarrollar el modelo matemático que describa el crecimiento del arándano Biloxi en función de las variables ambientales.
5. **Validación y ajuste del modelo:** Validar el modelo desarrollado y realizar los ajustes necesarios para mejorar su precisión y aplicabilidad, realizando los ajustes y pruebas necesarias del módulo web.
6. **Documentación y difusión de resultados:** Documentar los hallazgos y las conclusiones del estudio, y realizar despliegue del módulo web que muestre el modelo y los diferentes datos recolectados.

Resultados

Como resultado se tiene que el comportamiento del crecimiento vegetal conforme a una curva S o a una función sigmoide por lo que por medio del modelo comparativo y descriptivo se busca un ajuste o tener un punto de crecimiento “ideal” y así mismo poder intervenir según sea el caso si esta alta o baja la medición que brinden los sensores activar los actuadores para poder ajustar la variable que se esté midiendo en el momento, como se muestra en la figura 1.

Figura 1

Ejemplo de una curva de crecimiento ideal, o sea que coincide con el desarrollo en altura deseado y dos curvas de crecimiento que se tienden a alejar de la ideal por defecto y por exceso a partir de la fase de crecimiento inicial; estas tendencias se pueden corregir aumentando o disminuyendo la tasa de fertilizante respectivamente.



Nota. Fuente: tomada de Fases de cultivo: establecimiento y crecimiento rápido https://www.researchgate.net/publication/272790732_Fases_de_cultivo_establecimiento_y_crecimiento_rapido, 2012.

Por otro lado, se cuenta con diferentes modelos que permiten el desarrollo y comprensión del crecimiento resaltando entre estos los Modelos Aditivos Generalizados (GAM), los cuales en lugar de suponer una relación lineal entre las variables, los GAM permiten relaciones no lineales al introducir funciones suaves sobre las variables predictoras (Aguilar, 2019), también se tienen los modelos mecanicistas que son representación abstracta de un mecanismo que se divide en dos partes: una describe el comportamiento interno de los componentes del mecanismo y la otra describe el comportamiento global del mecanismo siendo estos modelos en compañía de los estadísticos la opción viable para poder enfocar la investigación a los resultados que se quieren llegar, para ir terminando la parte de los resultados se resalta que los modelos estadísticos incluyen diferentes técnicas que usan grandes bases de datos que, por lo general describen datos reales dentro de un sistema de forma empírica, siendo estos datos variables que interactúan y evolucionan entre sí en el tiempo (Bravo et al. 2019), cómo se plantea en diferentes estudios, los modelos lineales simples y multivariados son útiles para describir procesos que plantean una relación entre las variables predictoras y objetivo (Vásquez- Martínez et al., 2022; Kellam, 2023; Tasnim et al., 2022),

Así mismo, también se implementarán modelos descriptivos y comparativos para poder tener una idea mas centrada de lo que el modelo muestre en el modulo web para esto se tendrán en cuenta diferentes técnicas estadísticas como MANOVA, ANOVA, gráficos de pastel y dispersión y entre otras

Discusión

Los modelos mecanicistas y los modelos estadísticos representan dos enfoques distintos para describir y comparar el crecimiento de las plantas. Mientras que los modelos mecanicistas se basan en la comprensión de los procesos biológicos y físicos subyacentes, los modelos estadísticos se centran en el análisis de datos empíricos y la identificación de patrones y relaciones.

Los modelos mecanicistas, también conocidos como modelos de simulación o modelos de proceso, intentan representar los mecanismos fundamentales que rigen el crecimiento y desarrollo de las plantas (Craver, 2006). Estos modelos incorporan conocimientos teóricos y experimentales sobre la fisiología vegetal, los procesos bioquímicos y las interacciones con el entorno (Machamer et al., 2000). Al capturar estas interacciones complejas, los modelos mecanicistas pueden proporcionar una comprensión más profunda de cómo las variables ambientales influyen en el crecimiento de las plantas (Glennan, 2017). Sin embargo, el desarrollo de estos modelos puede ser desafiante debido a la complejidad inherente de los sistemas biológicos y la necesidad de datos detallados sobre los procesos internos.

Por otro lado, los modelos estadísticos se basan en el análisis de datos observados y la identificación de patrones y relaciones empíricas (Bravo et al., 2019). Estos modelos utilizan técnicas como la regresión, el análisis de varianza y los modelos aditivos generalizados (GAM) para relacionar las variables ambientales con las medidas de crecimiento de las plantas (Kisiel et al., 2023; Aguilar, 2019). Los modelos estadísticos pueden ser más flexibles y adaptables a diferentes conjuntos de datos, pero no necesariamente capturan los mecanismos biológicos subyacentes (Ruiz et al., 2011).

En el contexto de la descripción y comparación del crecimiento de las plantas, ambos enfoques pueden ofrecer información valiosa. Los modelos mecanicistas pueden proporcionar una comprensión más profunda de los procesos involucrados y permitir simulaciones más

precisas bajo diferentes condiciones ambientales (Hernandez et al., 2009). Por otro lado, los modelos estadísticos pueden ser más adecuados para analizar y comparar grandes conjuntos de datos empíricos, identificar patrones y relaciones complejas, y realizar predicciones basadas en observaciones pasadas (Domijan et al., 2006).

En muchos casos, una combinación de ambos enfoques puede ser beneficiosa. Los modelos mecanicistas pueden informar el desarrollo de modelos estadísticos más precisos, mientras que los modelos estadísticos pueden ayudar a identificar patrones y relaciones que puedan ser incorporados en modelos mecanicistas más sofisticados (Ritchie & Nesmith, 1991; Vásquez-Martínez et al., 2022).

Además, las técnicas estadísticas multivariantes, como el análisis de varianza multivariante (MANOVA), el análisis discriminante y el análisis de correlación canónica, pueden ser útiles para comparar el crecimiento de las plantas en diferentes sistemas de cultivo y analizar las relaciones entre múltiples variables ambientales y de crecimiento (Rencher & Christensen, 2022; Hair et al., 2019; Huberty & Olejnik, 2006)..

Conclusiones

- 1) Los resultados obtenidos permiten concluir de forma afirmativa que el análisis conjunto de datos empíricos y la comprensión de los mecanismos biológicos subyacentes son fundamentales para el desarrollo de modelos matemáticos precisos y confiables en el ámbito de la producción agrícola. La combinación de estos enfoques complementarios proporciona una visión integral del fenómeno estudiado y permite aprovechar las fortalezas de cada uno de ellos.
- 2) Se concluye que el modelo matemático desarrollado en este estudio representa una herramienta valiosa para la toma de decisiones informadas y la planificación estratégica en la producción de arándanos Biloxi. Al ser capaz de describir y comparar el comportamiento del cultivo bajo diferentes condiciones ambientales, el modelo puede contribuir a la mejora de las prácticas de cultivo, mejor aprovechamiento de los recursos tanto naturales como químicos, y el aumento de la productividad.

Referencias

- Aguilar Barreiro, P. (2019). Modelos aditivos generalizados. (Trabajo Fin de Grado Inédito). Universidad de Sevilla, Sevilla.
- Bravo, C., Quintana, J. R., & Fernández-Manjalres, A. J. (2019). Modelización del crecimiento y desarrollo de plantas de café basada en datos experimentales y relaciones fenológicas con variables climáticas. *Cenicafé*, 70(2), 58-72.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese*, 153(3), 355-376.
<https://doi.org/10.1007/s11229-006-9097-x>
- Domijan, D., Sigg, P., Hinz, U., & Sauter, U. (2006). Plant growth models with statistical explanatory variables. *Plant Biology*, 8(4), 500-507.
- Gibergans, Josep (2019). Regresión lineal múltiple. Universitat Oberta de Catalunya.
- Glennan, S. (2017). The new mechanical philosophy. Oxford University Press.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hernandez, J. A., Galindo, P., Navarro, J. A., & Pron, N. (2009). Model for the prediction of sugarcane crop ripening in Guaduas Valley (Colombia). *Ingeniería e Investigación*, 29(2), 49-57.
- Hilty, J., Muller, B., Pantin, F. and Leuzinger, S. (2021), Plant growth: the What, the How, and the Why. *New Phytol*, 232: 25-41. <https://doi.org/10.1111/nph.17610>.
- Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). Wiley.
- Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 183-208). Academic Press.
- Kellam, E. (2023). *Statistical modeling of biological systems*. Cambridge University Press.
- Kisiel, N., Witek, B. B., & Brocki, S. (2023). Statistical models for predicting the growth of ornamental plants. *Journal of Horticultural Science and Biotechnology*, 98(1), 1-10.

- Lejarza, A. (2011). Modelos lineales y no lineales en la modelización de datos. Universidad del País Vasco.
- Leandro, R., Guzmán, F., & Vargas, C. (2019). Técnicas descriptivas multivariantes: Una guía práctica para su aplicación. Ediciones UNAM.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67(1), 1–25. <http://www.jstor.org/stable/188611>
- Manzano, A., Jiménez, M., & Rodríguez, J. (2019). Modelos comparativos aplicados a la investigación en ciencias sociales. Editorial Académica.
- Rencher, A. C., & Christensen, W. F. (2022). *Methods of multivariate analysis* (4th ed.). Wiley.
- Ritchie, J. T., & Nesmith, D. S. (1991). Temperature and crop development. Modeling plant and soil systems (pp. 5-29). American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.
- Ruiz, G. A., Crossa, J., Rodríguez, F., Barón, R., & Castañón, G. (2011). Modelos lineales, no lineales y temporales para el pronóstico de rendimientos de frijol de temporal. *Agricultura Técnica en México*, 37(3), 293-305. <https://www.revistafitotecniamexicana.org/documentos/37-3/8a.pdf>
- Vásquez-Martínez, M., Lizarazo-Peña, P., Darghan, E., Moreno-Fonseca, L. P. and Magnitskiy, S. (2022). Leaf area prediction models from growth measurements in Andean blueberry (*Vaccinium meridionale Swartz*) in the nursery. *Agronomía Colombiana*, 40(3), 361–371. <https://doi.org/10.15446/agron.colomb.v40n3.105039>

Development of a mathematical model to describe biloxi blueberry growth by comparison of environmental variables

Desarrollo de un modelo matemático para describir el crecimiento de arándanos biloxi mediante comparación de variables ambientales

Oscar Javier Morera Zarate^{1, 2}, José Alfonso Olaya Benavides^{2, 3}, Ariel Camilo Sánchez López^{3, 4}, Jhonatan Murcia Posso⁴

¹ Universidad de Cundinamarca, Facultad de Ingeniería en Sistemas y Computación, Grupo de investigación GISTFA Colombia. E-mail: omorera@ucundinamarca.edu.co

² Universidad de Cundinamarca, Facultad de Ingeniería en Sistemas y Computación, Grupo de investigación AXIOMA Colombia. E-mail: jalfonsoolaya@ucundinamarca.edu.co

³ Universidad de Cundinamarca, Facultad de Ingeniería en Sistemas y Computación, Grupo de investigación GISTFA Colombia. E-mail: arielcsanchez@ucundinamarca.edu.co

⁴ Universidad de Cundinamarca, Facultad de Ingeniería en Sistemas y Computación, Grupo de investigación GISTFA Colombia. E-mail: jmurciap@ucundinamarca.edu.co

Abstract

This paper describes the process used to model plant growth using a mathematical model. The procedure began with the generation of synthetic data from real data, due to the low volume of the latter. Subsequently, the results obtained were trained and analyzed using a Generalized Additive Model (GAM), which made it possible to understand and determine the impact of each of the environmental variables evaluated

Key words: Growth, mathematical model, synthetic data, analysis.

Resumen

El presente artículo describe el proceso utilizado para modelar el crecimiento de una planta mediante un modelo matemático. El procedimiento comenzó con la generación de datos sintéticos a partir de datos reales, debido al bajo volumen de estos últimos. Posteriormente, se realizó el entrenamiento y análisis de los resultados obtenidos mediante un Modelo Aditivo Generalizado (GAM), lo cual permitió comprender y determinar el impacto de cada una de las variables ambientales evaluadas.

Palabras clave: Crecimiento, modelo matemático, datos sintéticos, análisis

1. INTRODUCCIÓN

El desarrollo de herramientas que permitan comprender de una forma más detallada el crecimiento de las plantas respecto a sus variables ambientales y de cómo estas tienen cierto impacto conforme pasa el tiempo, permite realizar estrategias para obtener un resultado adecuado, por ello se decidió realizar un modelo matemático que permitiera tener una comprensión del efecto que tienen las variables en el crecimiento de la planta.

El Modelo que se decidió implementar fue un modelo **GAM (Generalized Additive Model)** el cual se distingue por su capacidad para modelar relaciones no lineales entre variables predictoras y una variable de respuesta, manteniendo la interpretabilidad del modelo. A diferencia de los modelos lineales tradicionales, los **GAM** permiten que cada predictor influya en la variable de salida a través de una función suave no paramétrica, lo que proporciona mayor flexibilidad en el ajuste de datos complejos [1].

Esto permitió abordar rangos en donde las variables no tuvieran un aporte ni muy bajo ni muy alto, Una característica clave es que los **GAM** separan los efectos de cada predictor y los modelan aditivamente, lo cual es útil cuando se sospecha que la relación entre variables no es estrictamente lineal, pero tampoco se quiere imponer una forma funcional rígida [2], además, los **GAM** conservan la interpretabilidad, ya que cada término de la suma representa el efecto aislado de una variable, lo que permite visualizar y entender mejor la influencia de cada predictor [3].

Como última parte del modelo se trabajó también la correlación entre las variables esto para asegurarse de que no hubiera colinealidad respecto a las variables, y que estas no guardaban una relación fuerte unas con otras.

2. METODOLOGÍA

La investigación que se está realizando se presenta como un caso de estudio sobre el comportamiento y el crecimiento de arándanos biloxi en ambientes cerrados y abiertos de recirculación de agua dando paso a una investigación experimental de diferentes ambientes que permiten tener una relación de las diferentes variables y como estas afectan a la planta.

El diseño que se busca implementar es un modelo comparativo y descriptivo, ya que por un lado se comparan las variables ambientales y el crecimiento de las plantas en los diferentes ambientes, lo que implica también un enfoque comparativo buscando describir y modelar la relación entre estas variables. La población de estudio son 4-5 plantas de arándanos de variedad biloxi que se encuentran en espacios cerrados (invernadero y ambiente controlado) y abierto (ambiente normal) donde se recolectan datos de diferentes variables en tiempo real como de temperatura, humedad, luz y entre otras que son importantes para un buen desarrollo de la planta, para poder medir estos datos se implementan sensores y actuadores que por medio de técnicas de análisis permitan dar datos más limpios que serán entregados al modelo para ser procesados.

Por último, se describen las fases de estudio que son las siguientes

- **Revisión de literatura y planificación:** Realizar una revisión exhaustiva de la literatura relevante y planificar el diseño experimental, la recolección de datos y los análisis a realizar.
- **Preparación del entorno de cultivo:** Establecer los sistemas cerrados (invernadero y ambiente controlado) y abiertos (ambiente normal) para el cultivo de las plantas de arándano Biloxi.
- **Recolección de datos:** Monitorear y recolectar datos sobre las variables ambientales y el crecimiento de las plantas en los sistemas de cultivo durante un período de tiempo completo.
- **Análisis de datos:** Aplicar técnicas estadísticas y de modelado matemático para analizar los datos recolectados y desarrollar el modelo matemático que describa el crecimiento del arándano Biloxi en función de las variables ambientales.
- **Validación y ajuste del modelo:** Validar el modelo desarrollado y realizar los ajustes necesarios para mejorar su precisión y aplicabilidad.
- **Documentación y difusión de resultados:** Documentar los hallazgos y las conclusiones del estudio.

3. GENERACIÓN DE DATOS SINTÉTICOS PARA EL MODELO

El **prototipo computacional** para el manejo de las plantas cuenta con **3 ambientes** con diferentes características. **El primero** de estos ambientes es **semicontrolado** donde hay **3 plantas y 2 grupos de sensores**, grupos conformado por un **sensor capacitivo de humedad, una fotorresistencia, un DHT22 y un sensor de UV**. Este primer ambiente se considera **semicontrolado** porque a pesar de tener una fuente de luz artificial, también recibe luz solar y flujo de aire del exterior. **El segundo ambiente** es controlado, cuenta con **una planta y un grupo de sensores**. Este ambiente recibe luz solar en muy poca medida y se encuentra en su mayoría cerrado por lo que no hay mucho flujo de aire exterior. Y



finalmente el **tercer ambiente**, se encuentra abierto y con incidencia casi total de luz solar.

Para los procesos a continuación se tuvo en cuenta el **ambiente semicontrolado**, con información del ambiente mediante los **dos grupos de sensores y de la altura de las 3 plantas correspondientes a este ambiente**. Por razones externas al proyecto no fue posible contar con información de los demás ambientes, pero en un futuro esos datos estarán disponibles para nuevas investigaciones.

Con la intención de capturar la forma no lineal de los datos se pensó en hacer uso de una red neuronal, dada su capacidad para adaptarse a patrones propios de los datos [4]. Para esto se usó una red neuronal estructurada **con 3 capas ocultas de 64, 32 y 16 neuronas respectivamente usando la función de activación ReLU y una sola neurona de salida** para realizar la regresión del dato a predecir.

Adicionalmente, buscando ajustarse lo máximo posible a la forma de los datos se emplearon **300 épocas**, lo cual favoreció un ajuste más fino sobre la información real en comparación con otro modelo de menos épocas no presente en este documento.

Logrando obtener las representaciones para cada una de las variables **Figura 1, 2 y 3s**.

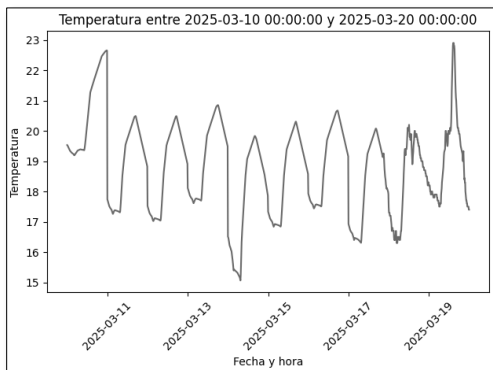


Figura 1. Datos sintéticos y reales de temperatura

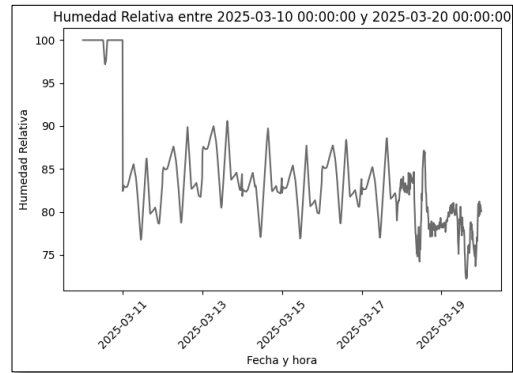


Figura 2. Datos sintéticos y reales de humedad de sustrato

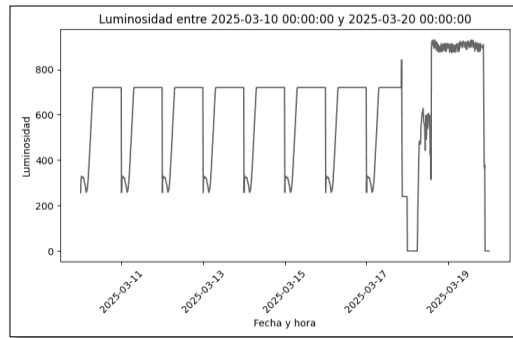


Figura 3. Datos sintéticos y reales de luminosidad

Todos estos procesos se replicaron para las demás variables de los ambientes que se tenían presentes

Variable dependiente altura: Respecto a esta variable se contaba con mediciones ocasionales, por esta razón, fue necesario también reconstruir la altura usando esos puntos en el tiempo, con esta información también se interpolaron datos para poder completar la serie temporal al igual que en las variables ambientales, dando como resultado la gráfica de la figura 4 que permite ver el crecimiento de las 3 plantas en el ambiente semicontrolado.

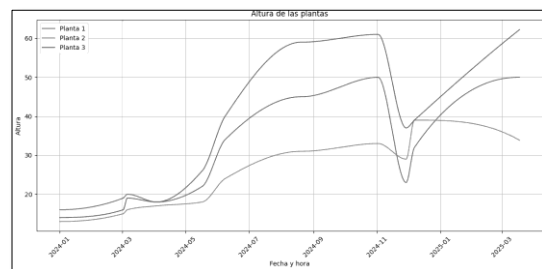


Figura 4. Altura de las 3 plantas del ambiente semicontrolado

4. RESULTADOS

Para poder lograr lo que se buscaba con el proyecto se optó por un modelo GAM, el cual permite modelar las relaciones no lineales entre las variables realizando funciones suavizadas [1], teniendo presente su formula la cual es la siguiente.

$$y_i = B_0 + \sum_{j=1}^p \left(\sum_{k=1}^{K_j} B_{jk} b_{jk}(x_{ij}) \right) \quad (1)$$

Donde

- y_i = Variable respuesta (Dependiente)
- B_0 = Terminio de intercepción
- p = Numero de predicciones
- K_j = Numero de funciones base asociados al predictor i
- B_{jk} = Coeficiente del modelo para función base b_{jk}
- $b_{jk}(x_{ij})$ = Funciones base evaluadas en el valor x_{ij}

Este enfoque presentado en la ecuación (1) permite capturar relaciones no lineales complejas sin necesidad de especificar una forma funcional rígida, y se emplea comúnmente en contexto donde requiere mayor flexibilidad que en la regresión lineal tradicional, pero con control sobre el sobreajuste mediante penalizaciones [2].

Para el desarrollo del modelo GAM se empleo como lenguaje de programación **Python**, con ayuda de diferentes librerías como **pandas** y **numpy** para el manejo de datos y arreglo de datos, **Pygam** para el manejo del modelo **GAM** esta librería ofrece diferentes variantes del modelo, por eso se emplea el módulo de **LinearGAM** (Modelo de regresión aditiva lineal), junto a sus parámetros que son importares para el desarrollo **S(Función para variables numéricas continuas (Splines))** y **F (Función para variables categóricas (factoriales))**.

Por otro lado, con ayuda de la librería de **Sklearn** y su modulo **Model_selection** se divide la cantidad de datos en entrenamiento y prueba con el módulo de **train_test_split**, para las métricas de evaluación se usó el módulo de **metrics** usando las siguientes métricas **r2_score**, **means_squared_error**, las gráficas con ayuda de **matplotlib**.

Estructura del modelo programado

Definimos X (predictores) y (variable objetivo)

```
X = df[['temp_high', 'temp_low', 'hum_rel', 'min_siembra', 'hum_sustr', 'lum', 'temp', 'poda_engrosamiento']]
y = df['altura']
```

División en entrenamiento y prueba (80% train, 20% test)

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42) # random_state para garantiza que los resultados sean reproducibles.
```

```
gam = LinearGAM(
    s(0, n_splines=10) +
    s(1, n_splines=10) +
    s(2, n_splines=10) +
    s(3, n_splines=10) +
    s(4, n_splines=10) +
    s(5, n_splines=10) +
    s(6, n_splines=10) +
    f(7)
).fit(X_train, y_train)
```

```
joblib.dump(gam, 'gam_model.pkl')
```

```
gam = joblib.load('gam_model.pkl')
```

Predicciones sobre conjunto de prueba

```
y_pred = gam.predict(X_test)
```

Métricas de evaluación

```
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
```

```
print(f"R2 (Porcentaje explicado por el modelo): {r2*100:.2f}%")
```

```
print(f"RMSE (Error cuadrático medio): {rmse:.4f}")
```

Definición de variables predictoras y objetivo

Como se muestra las variables predictoras corresponden a los datos recolectados por los sensores, tales como temperatura, humedad, luminosidad, tiempo de siembra, poda y entre otras, esto empleado para ver su influencia de sobre la variable objetivo que en este caso es la altura de la planta.

Tabla 1 Descripción de variables predictoras y categóricas

Spliner	Variable	Variable real
S(0)	temp_high	Temperatura alta
S(1)	temp_low	Temperatura mínima
S (2)	hum_rel	Humedad relativa en aire
S (3)	min_siembra	Tiempo transcurrido desde la siembra
S (4)	hum_sustr	Humedad de sustrato
S (5)	lum	Luminosidad
S (6)	temp	Temperatura media
F (7)	poda_engrosamiento	Poda engrosamiento (Variable categórica)

Entrenamiento del modelo: El conjunto de datos se divide en dos partes: el **80%** se utiliza para el entrenamiento del modelo y el **20%** restante para la prueba, empleando un valor de **random_state=42** para garantizar la reproducibilidad de los resultados.

Para el entrenamiento se utiliza el modelo **LinearGAM**, el cual permite sumar efectos no lineales de cada variable. Este modelo emplea distintos tipos de funciones según la naturaleza de los datos: **s(i)** para suavizar variables numéricas, **f(i)** para tratar variables categóricas y **fit()** para ajustar el modelo a los datos de entrenamiento.

Métricas de pruebas para el modelo GAM: Las métricas de pruebas que se emplean con ayuda de la librería **Sklearn** se usa el **R² (Coeficiente de determinación)** el cual mide que tan bien las predicciones explican la variabilidad de los datos reales y el **RMSE** el cual en cuanto menor sea mejor es, ya que mide cuanto error hay en promedio entre las predicciones y los valores reales

Predicciones sobre conjunto de prueba
`y_pred = gam.predict(X_test)`

Métricas de evaluación
`r2 = r2_score(y_test, y_pred)`

`rmse = np.sqrt(mean_squared_error(y_test, y_pred))`

Tabla 2 Resultado de métricas del modelo

R² (Porcentaje explicado por el modelo):	97.83%
RMSE (Error cuadrático medio):	1.8589

Como se muestran los resultados de las métricas de evaluación del modelo permite ver que el modelo da un buen resultado al ajustarse y predecir los valores basados en línea de curva central.

$$\text{altura} = s(0) + s(1) + s(2) + s(3) + s(4) + s(5) + s(6) + f(7) + \text{intercepto} \quad (2)$$

La ecuación **(2)** explica la estructura del modelo **GAM**, Lo que hace es sumar el efecto de cada función suavizada **spline S(i)** o variable predictora, la función lineal de la variable categórica **f(7)** la cual corresponde a la poda de engrosamiento expresada como una variable categórica y el intercepto (valor base).

Tabla 3. Aporte de las variables al modelo

Feature Funcion	Lambda	Ran k	EDo F	P > x	Sig. Code
s(0)	[0.6]	10	8.8	1.11e-16	***
s(1)	[0.6]	10	8.5	1.11e-16	***
s(2)	[0.6]	10	7.6	1.11e-16	***
s(3)	[0.6]	10	8.5	1.11e-16	***
s(4)	[0.6]	10	7.8	6.66e-16	***
s(5)	[0.6]	10	7.6	7.35e-01	
s(6)	[0.6]	10	7.0	3.43e-04	***
f(7)	[0.6]	2	1.0	1.11e-16	***

Análisis de efectos parciales de cada una de las variables: Como se trata de un modelo lineal, cada variable tiene un efecto proporcional sobre el crecimiento de la planta. Los valores con efecto menor a cero van a restar al resultado final, es decir,

van a tener un impacto negativo en el crecimiento. En cambio, entre más alto sea el valor del efecto (positivo), mayor será su contribución al crecimiento. De esta forma, el modelo suma o resta según el peso que tenga cada variable.

En este caso, las únicas variables reales medidas directamente son la temperatura máxima y la temperatura mínima sobre Facatativá, usándose como fuente de información verídica al modelo. Las demás variables utilizadas en el modelo fueron generadas de forma sintética, con el fin de complementar la información y evaluar su posible influencia en el crecimiento de la planta

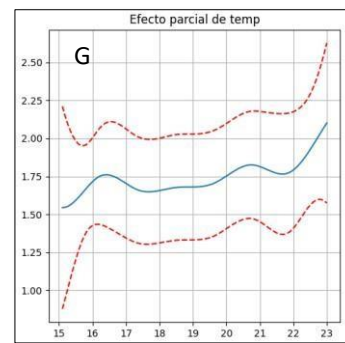
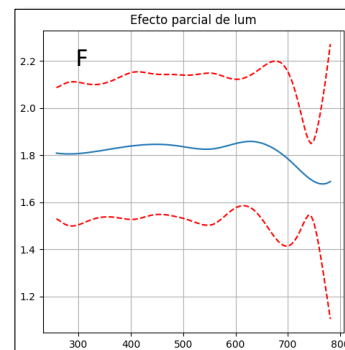
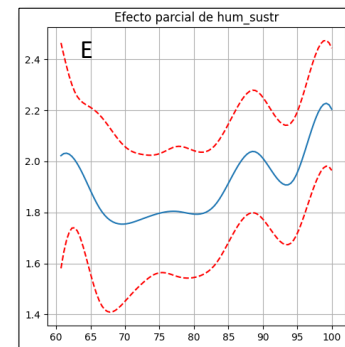
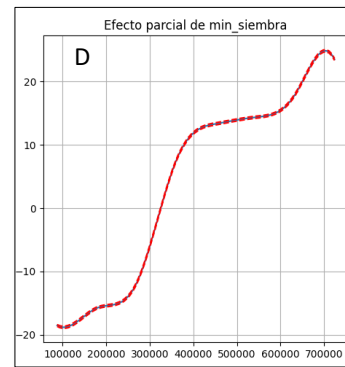
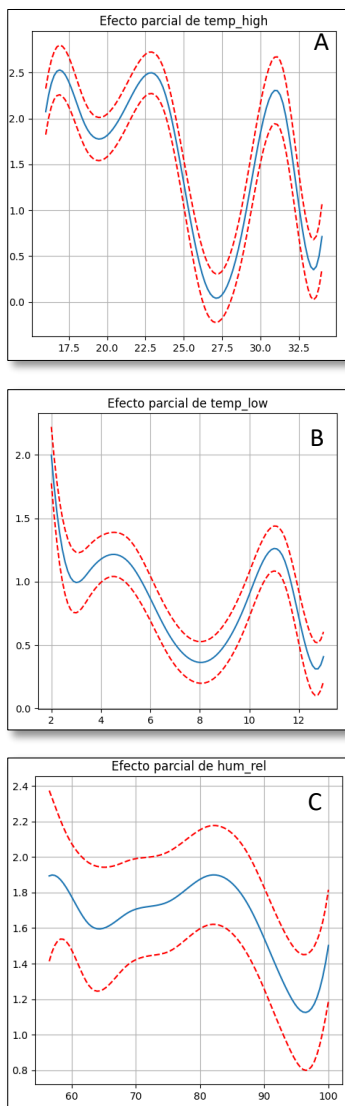


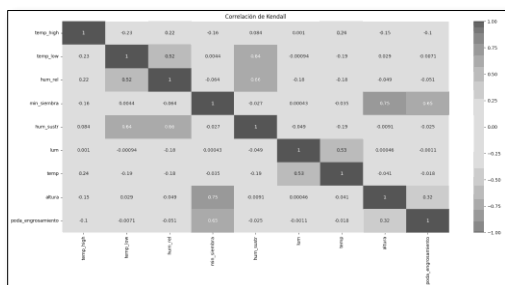
Figura 5. Graficas de efectos parciales. (A) temp_high, (B) temp_low, (C) hum_rel, (D) min_siembra, (E) hum_sustr, (F) lum, (G) temp

Estas graficas presentadas nos permiten tener un mejor entendimiento del comportamiento

individual de cada una de las variables respecto a la altura al pasar el tiempo, permitiendo obtener rangos en donde estas aporten de una manera más adecuada al crecimiento.

Por último, para respaldar los resultados obtenidos con el **GAM** se realizó una correlación de Kendall entre las variables la cual es una medida no paramétrica que evalúa la asociación entre dos variables ordinales, considerando la concordancia y discordancia entre pares de observaciones [5].

Figura 6. Matriz de correlación Kendall



Con esta matriz se puede inferir que no hay colinealidad significativa entre las variables evaluadas por el modelo, ya que la mayoría de los coeficientes se encuentran con valores bajos.

CONCLUSIONES

El modelo GAM, al ser una herramienta estadística flexible y robusta, permitió realizar un análisis detallado del efecto individual de cada variable sobre la altura de la planta. Las gráficas generadas demostraron su capacidad para capturar relaciones no lineales complejas, posicionándolo como una alternativa superior frente a modelos más rígidos. Además, el análisis de la correlación de Kendall entre las variables predictoras evidenció una baja multicolinealidad, lo cual favorece una interpretación más clara y precisa del impacto de cada predictor sobre la variable de respuesta. Finalmente, el uso del modelo GAM facilitó la identificación de los valores óptimos de las variables que promueven un mayor crecimiento de las plantas, aportando información valiosa para la toma de decisiones en contextos agronómicos y experimentales.

Los valores de rangos adecuados para estas variables se muestran en la siguiente tabla.

Tabla 4. Rangos con más aporte a la altura

Rangos	Variable	Variable real
18 a 23 grados Celsius	temp_high	Temperatura alta
2.6 a 6 grados Celsius	temp_low	Temperatura mínima
75 a 90	hum_rel	Humedad relativa en aire
Mayor de 85	hum_sustr	Humedad de sustrato
18 a 23 grados	lum	Luminosidad
	temp	Temperatura media

Estos valores obtenidos por medio del modelo identificando los rangos adecuados en que estos aportan a la altura de la planta, esto a raíz del análisis de las gráficas de efectos parciales de cada una de estas.

REFERENCIAS

- [1] T. Hastie and R. Tibshirani, Generalized Additive Models, Boca Raton: CRC Press, 1990. Chapter or Book
- [2] S. Wood, Generalized Additive Models: An Introduction with R, 2nd ed., Boca Raton: CRC Press, 2017. <https://doi.org/10.1201/9781315370279>
- [3] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed., New York: Springer, 2021.
- [4] E. Varela-Arregocés and E. Campbells-S, Redes Neuronales Artificiales: una revisión del estado del arte, aplicaciones y tendencias futuras, Universidad Simón Bolívar, Investigación y Desarrollo en TIC, 2011.
- [5] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1–2, pp. 81–93, Jun. 1938. [Online]. Available: <https://doi.org/10.1093/biomet/30.1-2.81>



UDEC
UNIVERSIDAD DE
CUNDINAMARCA

En el marco del Artículo 120 de la Ley 30 de 1992

Certifica que:

JHONATAN MURCIA POSSO

Identificado(a) con Cédula de ciudadanía N° 1006502314

Participó como **PONENTE** en la:

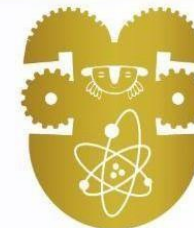
**V ENCUENTRO DE SEMILLEROS DE INVESTIGACIÓN: AVANCES Y
DESAFIOS DE LOS SEMILLEROS DE INVESTIGACIÓN DESDE LA
TRANSLOCALIDAD**

Organizada por el programa académico de Dirección de Investigación Sede Fusagasugá, la Facultad Dirección de Investigación, la Dirección de Investigación, la Dirección de Interacción Social Universitaria y la Vicerrectoría Académica, con una intensidad de 10 horas, realizado el día 22 de noviembre de 2023.

La presente certificación se expide a solicitud del(la) interesado(a) a los veintinueve (29) días del mes de abril de 2024.

ENA PATRICIA GIL BELLIDO
Directora Interacción Social Universitaria
Universidad de Cundinamarca

UNIVERSIDAD INCCA DE COLOMBIA



Reconocimiento como Universidad: Decreto Ejecutivo No. 687 del 6 de mayo de 1970
Personería Jurídica: Resolución Minjusticia No. 1891 del 19 de junio de 1963
Vigilada Mineducación

CERTIFICA QUE:

Jhonatan Murcia Posso

Identificado(a) con C.C. No. 1006502314

Participó como **Ponente** en el evento:

V Congreso Internacional de Ciencia, Tecnología E Innovación Para El Desarrollo Sustentable

Realizado el 22 y 23 de mayo del 2024

31 de mayo de 2024
Bogotá, D.C., Colombia

SANTIAGO SEPÚLVEDA CARDONA
VICERRECTOR GENERAL



LAURA MILENA VILLAR PIRAQUIVE
SECRETARIA GENERAL

“El presente documento se expide por única vez y certifica Educación Continua no formal, que no certifica aptitudes, saberes y no conduce a la obtención de un título”