	MACROPROCESO DE APOYO	CÓDIGO: AAAr113
	PROCESO GESTIÓN APOYO ACADÉMICO	VERSIÓN: 3
	DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL	VIGENCIA: 2017-11-16
		PAGINA: 1 de 7

16.

FECHA	viernes, 13 de noviembre de 2020
--------------	----------------------------------

Señores
UNIVERSIDAD DE CUNDINAMARCA
 BIBLIOTECA
 Ciudad


UNIDAD REGIONAL	Extensión Facatativá
TIPO DE DOCUMENTO	Trabajo De Grado
FACULTAD	Ingeniería
NIVEL ACADÉMICO DE FORMACIÓN O PROCESO	Pregrado
PROGRAMA ACADÉMICO	Ingeniería de Sistemas

El Autor(Es):

APELLIDOS COMPLETOS	NOMBRES COMPLETOS	No. DOCUMENTO DE IDENTIFICACIÓN
Viracacha Viracacha	Corin Jazmin	1070985185

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca
 Teléfono (091) 8281483 Línea Gratuita 018000976000
 www.ucundinamarca.edu.co E-mail: info@ucundinamarca.edu.co
 NIT: 890.680.062-2

*Documento controlado por el Sistema de Gestión de la Calidad
 Asegúrese que corresponde a la última versión consultando el Portal Institucional*

	MACROPROCESO DE APOYO	CÓDIGO: AAAR113
	PROCESO GESTIÓN APOYO ACADÉMICO	VERSIÓN: 3
	DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL	VIGENCIA: 2017-11-16
		PAGINA: 2 de 7

Director(Es) y/o Asesor(Es) del documento:

APELLIDOS COMPLETOS	NOMBRES COMPLETOS
Barahona Rodriguez	Cesar Yesid

TÍTULO DEL DOCUMENTO
APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE BRECHA DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL PAÍS

SUBTÍTULO (Aplica solo para Tesis, Artículos Científicos, Disertaciones, Objetos Virtuales de Aprendizaje)

TRABAJO PARA OPTAR AL TÍTULO DE: Aplica para Tesis/Trabajo de Grado/Pasantía
Ingeniero de sistemas

AÑO DE EDICION DEL DOCUMENTO	NÚMERO DE PÁGINAS
13/11/2020	251

DESCRIPTORES O PALABRAS CLAVES EN ESPAÑOL E INGLÉS (Usar 6 descriptores o palabras claves)	
ESPAÑOL	INGLÉS
1. Minería de datos	Data mining
2. Brechas de género	Gender gaps
3. Pruebas saber	Knowledge tests
4. Educación básica	Basic education
5. Clusterización	Clustering
6. Predicción	Prediction

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca
Teléfono (091) 8281483 Línea Gratuita 018000976000
www.ucundinamarca.edu.co E-mail: info@ucundinamarca.edu.co
NIT: 890.680.062-2

*Documento controlado por el Sistema de Gestión de la Calidad
Asegúrese que corresponde a la última versión consultando el Portal Institucional*



MACROPROCESO DE APOYO	CÓDIGO: AAAR113
PROCESO GESTIÓN APOYO ACADÉMICO	VERSIÓN: 3
DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL	VIGENCIA: 2017-11-16
	PAGINA: 3 de 7

RESUMEN DEL CONTENIDO EN ESPAÑOL E INGLÉS

(Máximo 250 palabras – 1530 caracteres, aplica para resumen en español):

Resumen

El presente proyecto está enfocado a determinar qué factores influyen en las brechas de género de la educación básica primaria y secundaria, para ello se tuvieron en cuenta los resultados académicos y socioeconómicos de las pruebas saber 3°, 5° y 9°; con esta información se realizó por medio de técnicas de minería de datos en la educación, un análisis en el cual se identifica la incidencia de cada variable sobre el comportamiento de los estudiantes. Para el análisis descriptivo se implementó el algoritmo K-Means, el cual se encarga de agrupar la información según sus características, de esta manera se identificaron los grupos con mayor concentración de datos; una vez obtenidos los resultados del algoritmo se demostró la relación de dichos resultados por año y tipo de prueba.

Asimismo se realizó un análisis predictivo, para determinar los posibles puntajes a obtener en matemáticas y lenguaje. Con la información socioeconómica de los estudiantes, se realizó un análisis por medio del algoritmo J48, el cual da como resultado un árbol de decisión; de este se extrajeron las reglas predictivas de los puntajes a obtener de acuerdo a cada una de las variables analizadas.

Con la información obtenida de los análisis descriptivos y predictivos se procedió a realizar un aplicativo, en el cual se muestra gráficamente los resultados y la relación de estos frente a cada año y tipo de prueba.

Abstract

This project is focused on determining which factors influence the gender gaps in primary education, for this purpose, the academic and socio-economic results of the tests were taken into account 3rd, 5th and 9th; with this information was done through data mining techniques in education, an analysis identifying the impact of each variable on student behavior. For descriptive analysis, the K-Means algorithm was implemented, which is responsible for grouping the information according to its characteristics, in this way the groups with the highest concentration of data were identified; once the results of the algorithm have been obtained, it seeks to demonstrate the relationship of these results by year and type of test. A predictive analysis was also conducted to determine the possible scores to be obtained in mathematics and language. With students' socioeconomic information, an analysis was performed using the J48 algorithm, which results in a decision tree, from this were extracted the predictive rules of the scores to be obtained according to each of the variables analyzed. With the information obtained from the descriptive and predictive analyses, an application was carried out, which graphically shows the results and their relationship to each year and type of test.



MACROPROCESO DE APOYO	CÓDIGO: AAAR113
PROCESO GESTIÓN APOYO ACADÉMICO	VERSIÓN: 3
DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL	VIGENCIA: 2017-11-16
	PAGINA: 4 de 7

AUTORIZACION DE PUBLICACIÓN

Por medio del presente escrito autorizo (Autorizamos) a la Universidad de Cundinamarca para que, en desarrollo de la presente licencia de uso parcial, pueda ejercer sobre mí (nuestra) obra las atribuciones que se indican a continuación, teniendo en cuenta que, en cualquier caso, la finalidad perseguida será facilitar, difundir y promover el aprendizaje, la enseñanza y la investigación.

En consecuencia, las atribuciones de usos temporales y parciales que por virtud de la presente licencia se autoriza a la Universidad de Cundinamarca, a los usuarios de la Biblioteca de la Universidad; así como a los usuarios de las redes, bases de datos y demás sitios web con los que la Universidad tenga perfeccionado una alianza, son:
Marque con una "X":

AUTORIZO (AUTORIZAMOS)	SI	NO
1. La reproducción por cualquier formato conocido o por conocer.	X	
2. La comunicación pública por cualquier procedimiento o medio físico o electrónico, así como su puesta a disposición en Internet.	X	
3. La inclusión en bases de datos y en sitios web sean éstos onerosos o gratuitos, existiendo con ellos previa alianza perfeccionada con la Universidad de Cundinamarca para efectos de satisfacer los fines previstos. En este evento, tales sitios y sus usuarios tendrán las mismas facultades que las aquí concedidas con las mismas limitaciones y condiciones.	X	
4. La inclusión en el Repositorio Institucional.	X	

De acuerdo con la naturaleza del uso concedido, la presente licencia parcial se otorga a título gratuito por el máximo tiempo legal colombiano, con el propósito de que en dicho lapso mi (nuestra) obra sea explotada en las condiciones aquí estipuladas y para los fines indicados, respetando siempre la titularidad de los derechos patrimoniales y morales correspondientes, de acuerdo con los usos honrados, de manera proporcional y justificada a la finalidad perseguida, sin ánimo de lucro ni de comercialización.

Para el caso de las Tesis, Trabajo de Grado o Pasantía, de manera complementaria, garantizo(garantizamos) en mi(nuestra) calidad de estudiante(s) y por ende autor(es) exclusivo(s), que la Tesis, Trabajo de Grado o Pasantía en cuestión, es producto de mi(nuestra) plena autoría, de mi(nuestro) esfuerzo personal intelectual, como consecuencia de mi(nuestra) creación original particular y, por tanto, soy(somos) el(los) único(s) titular(es) de la misma. Además, aseguro (aseguramos) que no



MACROPROCESO DE APOYO	CÓDIGO: AAAr113
PROCESO GESTIÓN APOYO ACADÉMICO	VERSIÓN: 3
DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL	VIGENCIA: 2017-11-16
	PAGINA: 5 de 7

contiene citas, ni transcripciones de otras obras protegidas, por fuera de los límites autorizados por la ley, según los usos honrados, y en proporción a los fines previstos; ni tampoco contempla declaraciones difamatorias contra terceros; respetando el derecho a la imagen, intimidad, buen nombre y demás derechos constitucionales. Adicionalmente, manifiesto (manifestamos) que no se incluyeron expresiones contrarias al orden público ni a las buenas costumbres. En consecuencia, la responsabilidad directa en la elaboración, presentación, investigación y, en general, contenidos de la Tesis o Trabajo de Grado es de mí (nuestra) competencia exclusiva, eximiendo de toda responsabilidad a la Universidad de Cundinamarca por tales aspectos.

Sin perjuicio de los usos y atribuciones otorgadas en virtud de este documento, continuaré (continuaremos) conservando los correspondientes derechos patrimoniales sin modificación o restricción alguna, puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación de los derechos patrimoniales derivados del régimen del Derecho de Autor.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “*Los derechos morales sobre el trabajo son propiedad de los autores*”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables. En consecuencia, la Universidad de Cundinamarca está en la obligación de RESPETARLOS Y HACERLOS RESPETAR, para lo cual tomará las medidas correspondientes para garantizar su observancia.

NOTA: (Para Tesis, Trabajo de Grado o Pasantía):

Información Confidencial:

Esta Tesis, Trabajo de Grado o Pasantía, contiene información privilegiada, estratégica, secreta, confidencial y demás similar, o hace parte de la investigación que se adelanta y cuyos resultados finales no se han publicado.

SI ___ NO X.

En caso afirmativo expresamente indicaré (indicaremos), en carta adjunta tal situación con el fin de que se mantenga la restricción de acceso.

LICENCIA DE PUBLICACIÓN

Como titular(es) del derecho de autor, confiero(erimos) a la Universidad de Cundinamarca una licencia no exclusiva, limitada y gratuita sobre la obra que se integrará en el Repositorio Institucional, que se ajusta a las siguientes características:



MACROPROCESO DE APOYO	CÓDIGO: AAAr113
PROCESO GESTIÓN APOYO ACADÉMICO	VERSIÓN: 3
DESCRIPCIÓN, AUTORIZACIÓN Y LICENCIA DEL REPOSITORIO INSTITUCIONAL	VIGENCIA: 2017-11-16
	PAGINA: 6 de 7

- a) Estará vigente a partir de la fecha de inclusión en el repositorio, por un plazo de 5 años, que serán prorrogables indefinidamente por el tiempo que dure el derecho patrimonial del autor. El autor podrá dar por terminada la licencia solicitándolo a la Universidad por escrito. (Para el caso de los Recursos Educativos Digitales, la Licencia de Publicación será permanente).
- b) Autoriza a la Universidad de Cundinamarca a publicar la obra en formato y/o soporte digital, conociendo que, dado que se publica en Internet, por este hecho circula con un alcance mundial.
- c) Los titulares aceptan que la autorización se hace a título gratuito, por lo tanto, renuncian a recibir beneficio alguno por la publicación, distribución, comunicación pública y cualquier otro uso que se haga en los términos de la presente licencia y de la licencia de uso con que se publica.
- d) El(Los) Autor(es), garantizo(amos) que el documento en cuestión, es producto de mi(nuestra) plena autoría, de mi(nuestro) esfuerzo personal intelectual, como consecuencia de mi (nuestra) creación original particular y, por tanto, soy(somos) el(los) único(s) titular(es) de la misma. Además, aseguro(aseguramos) que no contiene citas, ni transcripciones de otras obras protegidas, por fuera de los límites autorizados por la ley, según los usos honrados, y en proporción a los fines previstos; ni tampoco contempla declaraciones difamatorias contra terceros; respetando el derecho a la imagen, intimidad, buen nombre y demás derechos constitucionales. Adicionalmente, manifiesto (manifestamos) que no se incluyeron expresiones contrarias al orden público ni a las buenas costumbres. En consecuencia, la responsabilidad directa en la elaboración, presentación, investigación y, en general, contenidos es de mí (nuestro) competencia exclusiva, eximiendo de toda responsabilidad a la Universidad de Cundinamarca por tales aspectos.
- e) En todo caso la Universidad de Cundinamarca se compromete a indicar siempre la autoría incluyendo el nombre del autor y la fecha de publicación.
- f) Los titulares autorizan a la Universidad para incluir la obra en los índices y buscadores que estimen necesarios para promover su difusión.
- g) Los titulares aceptan que la Universidad de Cundinamarca pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.
- h) Los titulares autorizan que la obra sea puesta a disposición del público en los términos autorizados en los literales anteriores bajo los límites definidos por la universidad en el “Manual del Repositorio Institucional AAAM003”
- i) Para el caso de los Recursos Educativos Digitales producidos por la Oficina de Educación Virtual, sus contenidos de publicación se rigen bajo la Licencia Creative Commons: Atribución- No comercial- Compartir Igual.



j) Para el caso de los Artículos Científicos y Revistas, sus contenidos se rigen bajo la Licencia Creative Commons Atribución- No comercial- Sin derivar.



Nota:

Si el documento se basa en un trabajo que ha sido patrocinado o apoyado por una entidad, con excepción de Universidad de Cundinamarca, los autores garantizan que se ha cumplido con los derechos y obligaciones requeridos por el respectivo contrato o acuerdo.

La obra que se integrará en el Repositorio Institucional, está en el(los) siguiente(s) archivo(s).

Nombre completo del Archivo Incluida su Extensión (Ej. PerezJuan2017.pdf)	Tipo de documento (ej. Texto, imagen, video, etc.)
1. APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE BRECHA DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL PAÍS.pdf	Texto

En constancia de lo anterior, Firmo (amos) el presente documento:

APELLIDOS Y NOMBRES COMPLETOS	FIRMA (autógrafo)
Viracacha Viracacha Corin Jazmin	

12.1-51-20

**APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE
BRECHA DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL
PAÍS**

CORIN JAZMIN VIRACACHA VIRACACHA

**UNIVERSIDAD DE CUNDINAMARCA
FACULTAD DE INGENIERÍA
PROGRAMA INGENIERÍA DE SISTEMAS
FACATATIVÁ
2020**

**APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE
BRECHA DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL
PAÍS**

CORIN JAZMIN VIRACACHA VIRACACHA

DIRECTOR:

CESAR YESID BARAHONA RODRIGUEZ
Ingeniero en telecomunicaciones, Especialista de Gestión de Proyectos

**UNIVERSIDAD DE CUNDINAMARCA
FACULTAD DE INGENIERÍA
PROGRAMA INGENIERÍA DE SISTEMAS
FACATATIVÁ
2020**

AGRADECIMIENTOS

Inicialmente agradezco a mis padres y a mis hermanos por haberme apoyado durante el logro de mis metas y objetivos, también quiero agradecer al profesor Cesar Yesid Barahona quien con todo su conocimiento me guio durante la realización del proyecto de grado, adicionalmente por haber confiado y creído en mis habilidades y capacidades.

Agradezco igualmente a todos los docentes que me acompañaron durante toda esta de aprendizaje, llenaron de conocimientos y experiencias mi formación y por supuesto me impulsaron a llegar a este punto.

COMPROMISO DEL AUTOR

Yo, **Corín Jazmin Viracacha Viracacha** identificada con célula de ciudadanía No. **1070985185** y con cód. **461216279** 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



A handwritten signature in black ink, appearing to read 'Corín Jazmin Viracacha', is written over a horizontal line.

RESUMEN

El presente proyecto está enfocado a determinar qué factores influyen en las brechas de género de la educación básica primaria y secundaria, para ello se tuvieron en cuenta los resultados académicos y socioeconómicos de las pruebas saber 3°, 5° y 9°; con esta información se realizó por medio de técnicas de minería de datos en la educación, un análisis en el cual se identifica la incidencia de cada variable sobre el comportamiento de los estudiantes. Para el análisis descriptivo se implementó el algoritmo K-Means, el cual se encarga de agrupar la información según sus características, de esta manera se identificaron los grupos con mayor concentración de datos; una vez obtenidos los resultados del algoritmo se demostró la relación de dichos resultados por año y tipo de prueba.

Asimismo se realizó un análisis predictivo, para determinar los posibles puntajes a obtener en matemáticas y lenguaje. Con la información socioeconómica de los estudiantes, se realizó un análisis por medio del algoritmo J48, el cual da como resultado un árbol de decisión; de este se extrajeron las reglas predictivas de los puntajes a obtener de acuerdo a cada una de las variables analizadas.

Con la información obtenida de los análisis descriptivos y predictivos se procedió a realizar un aplicativo, en el cual se muestra gráficamente los resultados y la relación de estos frente a cada año y tipo de prueba.

Palabras clave

Minería de datos, brechas de género, pruebas saber, educación básica, clusterización, predicción

ABSTRACT

This project is focused on determining which factors influence the gender gaps in primary education, for this purpose, the academic and socio-economic results of the tests were taken into account 3rd, 5th and 9th; with this information was done through data mining techniques in education, an analysis identifying the impact of each variable on student behavior. For descriptive analysis, the K-Means algorithm was implemented, which is responsible for grouping the information according to its characteristics, in this way the groups with the highest concentration of data were identified; once the results of the algorithm have been obtained, it seeks to demonstrate the relationship of these results by year and type of test.

A predictive analysis was also conducted to determine the possible scores to be obtained in mathematics and language. With students' socioeconomic information, an analysis was performed using the J48 algorithm, which results in a decision tree, from this were extracted the predictive rules of the scores to be obtained according to each of the variables analyzed.

With the information obtained from the descriptive and predictive analyses, an application was carried out, which graphically shows the results and their relationship to each year and type of test.

Keywords

Data mining, gender gaps, knowledge tests, basic education, clustering, prediction

CONTENIDO

INTRODUCCIÓN	23
1. INFORME DE INVESTIGACIÓN	25
1.1. Estado del Arte	25
1.1.1. Minería de datos	26
1.1.2. Investigaciones a nivel internacional	29
1.1.3. Investigaciones a nivel nacional	34
1.2. Línea de Investigación	38
1.3. Planteamiento del problema y pregunta de investigación	38
1.4. Objetivo general y objetivos específicos	40
1.5. Alcance e impacto del proyecto	41
1.6. Metodología	42
1.7. Marcos de Referencia	43
1.7.1. Marco Teórico	43
1.7.2. Marco Legal	51
2. AUTOMATIZACIÓN DE DESCARGA DE PRUEBAS	54
2.1. Implementación de WinSCP	54
2.2. Configuración de AutoFTP Manager	58
3. SELECCIÓN DE INDICADORES	61
3.1. Configuración de WEKA	61
3.2. Selección de indicadores	65
4. ALGORITMO DE ANALISIS PREDICTIVO	66
4.1. Validación algoritmo predictivo	68
4.1.1. Prueba con indicador de consola de videojuegos	69
4.1.2. Prueba con indicador educación de la madre	72
4.1.3. Prueba con indicador de internet	74
5. DOCUMENTACIÓN DEL SOFTWARE	76
5.1. Plan del proyecto	76
5.2. Determinación de requerimientos	78
5.2.1. Modificación de requerimientos	83
5.3. Especificación del diseño	84
5.3.1. Modelo Entidad Relación	84
5.3.2. Roles	87
5.3.3. Diagramas de casos de uso	88
5.3.4. Diagramas de secuencia	97
5.3.5. Diagramas de actividad	119
5.3.6. Diagrama de clases	145
5.4. Diseños de los casos de prueba (CALISOFT)	147
5.4.1. Resultados de Calisoft	148
5.5. Diseños de los casos de prueba (SONARQUBE)	153

5.5.1. Resultados de SonarQube _____	154
5.6. Estimación de recursos _____	156
5.6.1. Detalles de rubros _____	157
6. RESULTADOS _____	160
6.1. SABER 3 _____	160
6.1.1. Puntaje en lenguaje _____	160
6.1.2. Puntaje en matemáticas _____	160
6.2. SABER 5 _____	161
6.2.1. Educación padre _____	161
6.2.2. Educación madre _____	161
6.2.3. Acceso a computador _____	162
6.2.4. Acceso a internet _____	163
6.2.5. Acceso a consola de videojuegos _____	163
6.2.6. Resultados puntaje en lenguaje _____	164
6.2.7. Resultados puntaje en matemáticas _____	165
6.3. SABER 9 _____	165
6.3.1. Educación del padre _____	165
6.3.2. Educación de la madre _____	166
6.3.3. Acceso a computador _____	166
6.3.4. Acceso a internet _____	167
6.3.5. Acceso a consola de video juegos _____	168
6.3.6. Resultado de puntaje en lenguaje _____	168
6.3.7. Resultado de puntaje en matemáticas _____	169
CONCLUSIONES Y RECOMENDACIONES _____	170
BIBLIOGRAFÍA _____	172

LISTA DE TABLAS

<i>Tabla 1.</i> Análisis por género de las pruebas PISA 2012	25
<i>Tabla 2.</i> Validación de datos con 77 atributos	31
<i>Tabla 3.</i> Validación de datos con 15 atributos	31
<i>Tabla 4.</i> Estudios relacionados con el rendimiento académico	37
<i>Tabla 5.</i> Cantidad de variables analizadas por prueba	66
<i>Tabla 6.</i> Descripción del Modelo entidad relación de la base de datos ingapge....	85
<i>Tabla 7.</i> Descripción del modelo entidad relación de la base de datos resultados_basico	86
<i>Tabla 8.</i> Descripción de caso de uso inicio de sesión.....	88
<i>Tabla 9.</i> Descripción de caso de uso visualización de los resultados.....	89
<i>Tabla 10.</i> Descripción de caso de uso consultar resultados de los indicadores	89
<i>Tabla 11.</i> Descripción de caso de uso consultar resultados del análisis predictivo	89
<i>Tabla 12.</i> Descripción de caso de uso reportes	89
<i>Tabla 13.</i> Descripción de caso de uso consultar registro de mujeres por prueba..	90
<i>Tabla 14.</i> Descripción de caso de uso cerrar sesión	90
<i>Tabla 15.</i> Descripción de caso de uso inicio de sesión.....	91
<i>Tabla 16.</i> Descripción de caso de uso cerrar sesión	92
<i>Tabla 17.</i> Descripción de caso de uso CRUD de usuarios	92
<i>Tabla 18.</i> Descripción de caso de uso creación usuarios	92
<i>Tabla 19.</i> Descripción de caso de uso consulta usuarios	92
<i>Tabla 20.</i> Descripción de caso de uso modificación usuarios.....	93
<i>Tabla 21.</i> Descripción de caso de uso eliminación usuarios.....	93
<i>Tabla 22.</i> Descripción de caso de uso cargar pruebas	93
<i>Tabla 23.</i> Descripción de caso de uso consultar listado de pruebas	93
<i>Tabla 24.</i> Descripción de caso de uso consultar cantidad de registros por prueba	94
<i>Tabla 25.</i> Descripción de caso de uso carga de los datos.....	95
<i>Tabla 26.</i> Descripción de caso de uso limpieza de los datos.....	95
<i>Tabla 27.</i> Descripción de caso de uso cálculo de los indicadores de brecha de género.....	96
<i>Tabla 28.</i> Descripción de caso de uso análisis predictivo	96
<i>Tabla 29.</i> Descripción de caso de uso extracción de los datos	97
<i>Tabla 30.</i> Descripción diagrama de secuencia: inicio de sesión	98
<i>Tabla 31.</i> Descripción diagrama de secuencia: cerrar sesión.....	99
<i>Tabla 32.</i> Descripción diagrama de secuencia: creación usuarios	100
<i>Tabla 33.</i> Descripción diagrama de secuencia: consulta usuarios.....	101
<i>Tabla 34.</i> Descripción diagrama de secuencia: modificación usuarios	103
<i>Tabla 35.</i> Descripción diagrama de secuencia: eliminación usuarios	104
<i>Tabla 36.</i> Descripción diagrama de secuencia: cargar pruebas	106
<i>Tabla 37.</i> Descripción diagrama de secuencia: extracción de los datos.....	107

<i>Tabla 38.</i> Descripción diagrama de secuencia: carga de los datos	108
<i>Tabla 39.</i> Descripción diagrama de secuencia: limpieza de los datos	109
<i>Tabla 40.</i> Descripción diagrama de secuencia: consultar listado de pruebas.....	109
<i>Tabla 41.</i> Descripción diagrama de secuencia: cálculo de los indicadores de brecha de género.....	111
<i>Tabla 42.</i> Descripción diagrama de secuencia: análisis predictivo	112
<i>Tabla 43.</i> Descripción diagrama de secuencia: consultar resultados de los indicadores.....	113
<i>Tabla 44.</i> Descripción diagrama de secuencia: consultar resultados del análisis predictivo	115
<i>Tabla 45.</i> Descripción diagrama de secuencia: reportes	116
<i>Tabla 46.</i> Descripción diagrama de secuencia: consultar registro de mujeres por prueba.....	117
<i>Tabla 47.</i> Descripción diagrama de secuencia: consultar cantidad de registros por prueba.....	118
<i>Tabla 48.</i> Descripción diagrama de actividad: inicio de sesión	120
<i>Tabla 49.</i> Descripción diagrama de actividad: cerrar sesión.....	121
<i>Tabla 50.</i> Descripción diagrama de actividad: creación usuarios	123
<i>Tabla 51.</i> Descripción diagrama de actividad: modificación usuarios	125
<i>Tabla 52.</i> Descripción diagrama de actividad: consulta usuarios.....	126
<i>Tabla 53.</i> Descripción diagrama de actividad: eliminación usuarios	127
<i>Tabla 54.</i> Descripción diagrama de actividad: cargar pruebas	129
<i>Tabla 55.</i> Descripción diagrama de actividad: extracción de los datos.....	130
<i>Tabla 56.</i> Descripción diagrama de actividad: carga de los datos	132
<i>Tabla 57.</i> Descripción diagrama de actividad: limpieza de los datos	133
<i>Tabla 58.</i> Descripción diagrama de actividad: consultar listado de pruebas.....	134
<i>Tabla 59.</i> Descripción diagrama de actividad: cálculo de los indicadores de brecha de género.....	136
<i>Tabla 60.</i> Descripción diagrama de actividad: consultar resultados de los indicadores.....	137
<i>Tabla 61.</i> Descripción diagrama de actividad: análisis predictivo	139
<i>Tabla 62.</i> Descripción diagrama de actividad: consultar resultados del análisis predictivo	141
<i>Tabla 63.</i> Descripción diagrama de actividad: reportes	142
<i>Tabla 64.</i> Descripción diagrama de actividad: consultar registro de mujeres por prueba.....	143
<i>Tabla 65.</i> Descripción diagrama de actividad: consultar cantidad de registros por prueba.....	144
<i>Tabla 66.</i> Especificación de diagrama de clases	146
<i>Tabla 67.</i> Caso de prueba: Validación de nomenclatura de base de datos	147
<i>Tabla 68.</i> Caso de prueba: Verificación de la estructura del modelado	147
<i>Tabla 69.</i> Caso de prueba: Complejidad del código	153
<i>Tabla 70.</i> Caso de prueba: Duplicidad de código	153
<i>Tabla 71.</i> Caso de prueba: Calidad del software	154

<i>Tabla 72.</i> Resumen de costos	156
<i>Tabla 73.</i> Detalle Personal.....	157
<i>Tabla 74.</i> Descripción de equipos.....	157
<i>Tabla 75.</i> Descripción de materiales e insumos.....	158
<i>Tabla 76.</i> Descripción de servicios tecnológicos.....	158
<i>Tabla 77.</i> Descripción de viajes	159
<i>Tabla 78.</i> Descripción de otros	159

LISTA DE FIGURAS

<i>Figura 1.</i> Etapas de KDD	27
<i>Figura 2.</i> Fases Metodología CRISP-DM.....	28
<i>Figura 3.</i> ¿Qué metodología usar para minería de datos?	29
<i>Figura 4.</i> Exactitud de los resultados según los atributos usados	33
<i>Figura 5.</i> Perfiles de estudiantes.....	36
<i>Figura 6.</i> Indicadores de brechas de género	44
<i>Figura 7.</i> Esquema de un árbol de decisión.....	49
<i>Figura 8.</i> SCRIPT de conexión al FTP.....	55
<i>Figura 9.</i> Archivo BAT para localizar script	55
<i>Figura 10.</i> Configuración de tarea programada	56
<i>Figura 11.</i> Configuración de tiempos de la tarea programada	57
<i>Figura 12.</i> Configuración de la ruta del “bat”.....	57
<i>Figura 13.</i> Ejecución de tarea programada.....	58
<i>Figura 14.</i> Configuración de perfil de conexión.....	59
<i>Figura 15.</i> Configuración de perfil de conexión.....	59
<i>Figura 16.</i> Configuración de ajustes de transferencia.....	60
<i>Figura 17.</i> Configuración de programación de tiempos	60
<i>Figura 18.</i> Configuración del archivo “DatabaseUtils.props”	61
<i>Figura 19.</i> Interfaz inicial de weka.....	62
<i>Figura 20.</i> Configuración de usuario y contraseña.....	62
<i>Figura 21.</i> Query de consulta.....	63
<i>Figura 22.</i> Selección del algoritmo.....	63
<i>Figura 23.</i> Configuración del algoritmo K-means	64
<i>Figura 24.</i> Ejecución del algoritmo K-Means	64
<i>Figura 25.</i> Ejemplo de clasificación “Educacion madre”	65
<i>Figura 26.</i> Clusters de “Educacion madre”	65
<i>Figura 27.</i> Selección del algoritmo J48.....	67
<i>Figura 28.</i> Árbol generado con algoritmo J48	68
<i>Figura 29.</i> Porcentaje de clasificación correcta	69
<i>Figura 30.</i> Porcentaje de clasificación incorrecta.....	70
<i>Figura 31.</i> Error absoluto medio.....	71
<i>Figura 32.</i> Porcentaje de clasificación correcta	72
<i>Figura 33.</i> Porcentaje de clasificación incorrecta.....	73
<i>Figura 34.</i> Error absoluto medio.....	73
<i>Figura 35.</i> Porcentaje de clasificación correcta	74
<i>Figura 36.</i> Porcentaje de clasificación incorrecta.....	75
<i>Figura 37.</i> Error absoluto medio.....	75
<i>Figura 38.</i> Cronograma de desarrollo de proyecto.....	77
<i>Figura 39.</i> Modelo Entidad Relación, Base de datos Ingapge	84
<i>Figura 40.</i> Modelo Entidad Relación Base de datos resultados basico	85
<i>Figura 41.</i> Casos de uso: Usuario.....	88

<i>Figura 42.</i> Casos de uso: Administrador.....	91
<i>Figura 43.</i> Casos de uso: base de datos	94
<i>Figura 44.</i> Casos de uso: Aplicativo.....	95
<i>Figura 45.</i> Casos de uso: Sistema	96
<i>Figura 46.</i> Diagrama de secuencia: inicio de sesión.....	97
<i>Figura 47.</i> Diagrama de secuencia: cerrar Sesión	98
<i>Figura 48.</i> Diagrama de secuencia: creación usuarios	99
<i>Figura 49.</i> Diagrama de secuencia: consulta usuarios	101
<i>Figura 50.</i> Diagrama de secuencia: modificación usuarios.....	102
<i>Figura 51.</i> Diagrama de secuencia: eliminación usuarios.....	104
<i>Figura 52.</i> Diagrama de secuencia: cargar pruebas	105
<i>Figura 53.</i> Diagrama de secuencia: extracción de los datos.....	106
<i>Figura 54.</i> Diagrama de secuencia: carga de los datos	107
<i>Figura 55.</i> Diagrama de secuencia: limpieza de los datos.....	108
<i>Figura 56.</i> Diagrama de secuencia: consultar listado de pruebas.....	109
<i>Figura 57.</i> Diagrama de secuencia: cálculo de los indicadores de brecha de género	110
<i>Figura 58.</i> Diagrama de secuencia: análisis predictivo	111
<i>Figura 59.</i> Diagrama de secuencia: consultar resultados de los indicadores	112
<i>Figura 60.</i> Diagrama de secuencia: consultar resultados del análisis predictivo ..	114
<i>Figura 61.</i> Diagrama de secuencia: reportes	116
<i>Figura 62.</i> Diagrama de secuencia: consultar registro de mujeres por prueba ...	117
<i>Figura 63.</i> Diagrama de secuencia: consultar cantidad de registros por prueba ..	118
<i>Figura 64.</i> Diagrama de actividad: inicio de sesión.....	119
<i>Figura 65.</i> Diagrama de actividad: cerrar sesión.....	121
<i>Figura 66.</i> Diagrama de actividad: creación usuarios	122
<i>Figura 67.</i> Diagrama de actividad: modificación usuarios.....	124
<i>Figura 68.</i> Diagrama de actividad: consulta usuarios	126
<i>Figura 69.</i> Diagrama de actividad: eliminación usuarios.....	127
<i>Figura 70.</i> Diagrama de actividad: cargar pruebas	128
<i>Figura 71.</i> Diagrama de actividad: extracción de los datos.....	130
<i>Figura 72.</i> Diagrama de actividad: carga de los datos.....	131
<i>Figura 73.</i> Diagrama de actividad: limpieza de los datos.....	133
<i>Figura 74.</i> Diagrama de actividad: consultar listado de pruebas	134
<i>Figura 75.</i> Diagrama de actividad: cálculo de los indicadores de brecha de género	135
<i>Figura 76.</i> Diagrama de actividad: consultar resultados de los indicadores	136
<i>Figura 77.</i> Diagrama de actividad: análisis predictivo	138
<i>Figura 78.</i> Diagrama de actividad: consultar resultados del análisis predictivo ...	140
<i>Figura 79.</i> Diagrama de actividad: reportes	142
<i>Figura 80.</i> Diagrama de actividad: consultar registro de mujeres por prueba	143
<i>Figura 81.</i> Diagrama de actividad: consultar cantidad de registros por prueba ...	144
<i>Figura 82.</i> Diagrama de clases	145
<i>Figura 83.</i> Resultados de evaluación de base de datos	148

<i>Figura 84.</i> Resultados de evaluación de los diagramas de casos de uso.....	149
<i>Figura 85.</i> Resultados de evaluación de los diagramas de secuencia.....	150
<i>Figura 86.</i> Resultados de evaluación de los diagramas de actividades.....	151
<i>Figura 87.</i> Resultados de evaluación del MER	152
<i>Figura 88.</i> Resultados de evaluación del diagrama de clases	152
<i>Figura 89.</i> Métricas a evaluar del proyecto	154
<i>Figura 90.</i> Complejidad del proyecto	154
<i>Figura 91.</i> Duplicidad del código.....	155
<i>Figura 92.</i> Calidad del software	155
<i>Figura 93.</i> Resultados puntaje en lenguaje saber 3°	160
<i>Figura 94.</i> Resultados puntaje en matemáticas saber 3°.....	160
<i>Figura 95.</i> Resultados educación del padre en pruebas saber 5°	161
<i>Figura 96.</i> Resultados educación de la madre en pruebas saber 5°	161
<i>Figura 97.</i> Resultados acceso a computador en pruebas saber 5°	162
<i>Figura 98.</i> Resultados acceso a internet en pruebas saber 5°	163
<i>Figura 99.</i> Resultados acceso a consola de videojuegos en pruebas saber 5° ...	163
<i>Figura 100.</i> Resultados puntaje en lenguaje en pruebas saber 5°	164
<i>Figura 101.</i> Resultados puntaje en matemáticas en pruebas saber 5°.....	165
<i>Figura 102.</i> Resultados educacion del padre en pruebas saber 9°	165
<i>Figura 103.</i> Resultados educación de la madre en pruebas saber 9°.....	166
<i>Figura 104.</i> Resultados de acceso a computador en pruebas saber 9°.....	166
<i>Figura 105.</i> Resultados de acceso a internet en pruebas saber 9°	167
<i>Figura 106.</i> Resultados de acceso a consola de videojuegos en pruebas saber 9°	168
<i>Figura 107.</i> Resultados puntaje en lenguaje en pruebas saber 9°	168
<i>Figura 108.</i> Resultados puntaje en matemáticas en pruebas saber 9°.....	169

LISTA DE ANEXOS

Anexo A Manual de usuarios.....	175
Anexo B Manual técnico.....	207
Anexo C Artículo CACIED.....	233
Anexo D Artículo CIETA.....	239
Anexo E Artículo CICI 2020.....	244
Anexo F Control y seguimiento.....	250

INTRODUCCIÓN

Las brechas de género representan una de las diferencias socioeconómicas de un país, atribuidas a acontecimientos históricos en donde se obstaculiza el ejercicio equitativo de los derechos tanto para hombres como para mujeres, desde la Organización de las Naciones Unidas se ha buscado promover la igualdad de género y el empoderamiento de la mujer en todos los sectores sociales; ya que en los últimos años se han presentado sucesos en donde la mujer labora más y su remuneración es inferior, así mismo se evidencia la baja participación en las carreras del STEM (ciencia, tecnología, ingeniería y matemáticas) y en los procesos de investigación científica.

Para el año 2019, Colombia ocupó el puesto 22 en el reporte de brechas de equidad de género del Foro Económico Mundial, lo cual es de gran preocupación para el país, ya que esta problemática afecta el crecimiento social, económico y científico. Por tal motivo es importante identificar desde los procesos de formación de los jóvenes que factores influyen y como a partir de ellos se pueden generar estrategias que mitiguen dicha problemática.

Es así como la educación juega uno de los papeles más importantes en la erradicación de dicha problemática, ya que existen actualmente diferencias sociales entorno a los métodos de enseñanza y aprendizaje de los niños, las cuales se ven influenciadas por el entorno familiar y social en el que viven. Por lo tanto, se ve necesario identificar inicialmente los factores que influyen, ya que para diversos autores existe una relación entre el desempeño académico, el entorno familiar, los atributos sociodemográficos y socioeconómicos de los estudiantes. Es importante recalcar que estos paradigmas sociales sobre el género débil se han repercutido

desde la etapa de la niñez, puesto que desde esta se le ha encomendado a las niñas las tareas del hogar y a los niños las tareas productivas.

Por tal motivo las pruebas ICFES en sus resultados no solo buscan determinar puntajes, sino examinar como estos se ven afectados por factores externos e influyen en los procesos de formación; es por ello que las pruebas saber son de carácter censal, lo cual permiten estimar mejor los parámetros poblacionales, también cuentan con información regional esto con el fin de establecer diferencias de género y por último son un instrumento de medición de las competencias de los estudiantes (Abadia y Bernal, 2016).

El presente proyecto busca identificar por medio de técnicas de minería de datos educativa que factores o indicadores influyen en el crecimiento de las brechas de género en la educación básica primaria y secundaria del país, asimismo establecer la influencia de dichos indicadores en los resultados de los estudiantes. Para mostrar los resultados obtenidos se realiza un aplicativo web desarrollado en JAVA, en donde se evidencia gráficamente cada uno de los análisis hechos.

1. INFORME DE INVESTIGACIÓN

1.1. Estado del Arte

Según estudios realizados por el Ministerio de Educación en las pruebas PISA (2012) se demuestra que en Colombia los varones superan a las niñas en matemáticas y ciencias mientras que ellas obtienen mejores resultados en lectura (Figura 1), de acuerdo a lo anterior se concluye que la desigualdad de género aumenta con el paso de los años, producto de la baja autoestima, los estereotipos de género, la violencia y por supuesto las expectativas de los padres.

Tabla 1. Análisis por género de las pruebas PISA 2012

Prueba	Hombres	Mujeres	Hombres – Mujeres	Situación de Colombia entre los 65 países
Matemáticas	390	365	25	Valor máximo
Lectura	394	412	-18	La 2ª más baja
Ciencias	408	390	18	Valor máximo

Fuente: (Mineducación, 2013)

Por otro lado, Valle, Fonnegra, Byron y Valle (2016) corroboran la relación del desempeño académico en relación con el entorno familiar, los atributos sociodemográficos y socioeconómicos de los estudiantes, para ello implementaron un modelo multinivel teniendo en cuenta los resultados de las pruebas TIMSS 2007, el cual indicó que el desempeño de las mujeres en el área de matemáticas es muy bajo con respecto al promedio de los países de la OCDE, mientras que el de los hombres alcanzan los logros más sobresalientes en la prueba, concluyendo que la brecha de género en instituciones oficiales es de 13.3 puntos con respecto al desempeño académico. Dentro de los análisis descriptivos se evidencia el bajo nivel de autoconfianza de las mujeres, concluyendo por parte de ellas que no son buenas

en matemáticas y en consecuencia disminuyendo los niveles de esfuerzo por esta área.

Es así como Abadía y Bernal (2016) indican que la diferencia en áreas como matemáticas y ciencias por parte de las niñas inciden tanto en el acceso a la educación superior como en la innovación, creatividad y competitividad del país, es por ello, que para determinar los factores socioeconómicos, generales y familiares que influyen en esta problemática, se usan los resultados de las pruebas Saber ya que:

1) son de carácter censal, lo que permite reducir los errores de estimación de los parámetros poblacionales; 2) tiene información regional, la cual puede ayudar a explicar las diferencias de género observadas; 3) pueden ser un mejor instrumento de medición de las competencias de los estudiantes, porque estos pueden estar motivados a hacer su mejor esfuerzo, dada la importancia de la prueba para el ingreso a la educación superior. (Abadía & Bernal, 2016, p.8)

1.1.1. Minería de datos

“La minería de datos es considerada mecanismo de explotación que consiste en la búsqueda y el descubrimiento de información valiosa dentro de grandes volúmenes de datos” (Ballesteros et al., 2013). Asimismo, Ballesteros, Sánchez-Guzmán y García (2013) indican que el análisis de los datos dentro de la minería de datos de la educación EDM, permite descubrir patrones que caractericen el comportamiento de los estudiantes según sus evaluaciones; a través del uso de modelos predictivos de forma cualitativa y cuantitativa. A partir de ello Ballesteros, Sanchez y Garcia usaron dentro de sus investigaciones KDD (Descubrimiento de Conocimiento de Bases de Datos) como proceso para la obtención de resultados acorde a los datos

a tratar; este proceso se divide en una serie de pasos (Figura 2): en primera instancia, el pre-procesamiento en esta etapa se realiza la selección, limpieza, preparación y transformación de los datos; en segunda instancia, búsqueda e identificación de patrones aquí se selecciona y aplica el método o modelo de DM más apropiado y en última instancia la evaluación, donde se valida la calidez de los resultados obtenidos.

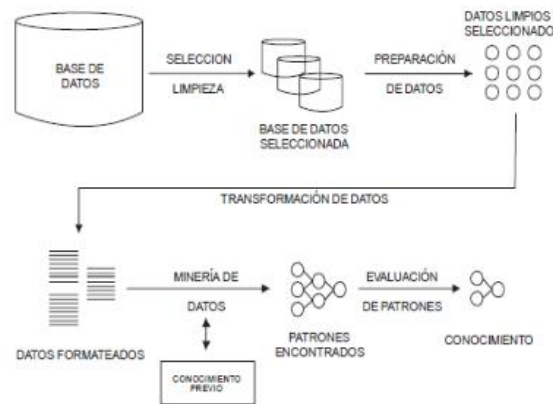


Figura 1. Etapas de KDD
Fuente: (Ballesteros et al., 2013)

Por otro lado CRISP-DM (Cross-Industry Standard Process for Data Mining) es considerada la “metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico e industrial” (Timaran & Jiménez, 2014), estructurada en seis etapas algunas bidireccionales, es decir que permite verificar las etapas anteriores (Figura 3).

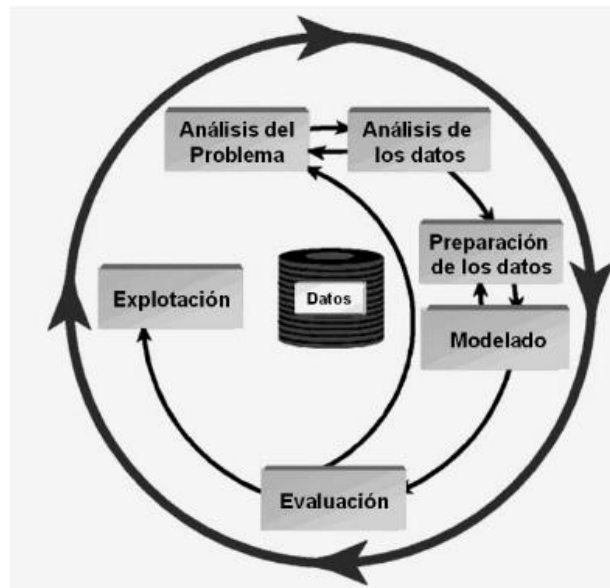


Figura 2. Fases Metodología CRISP-DM
Fuente: (Timaran & Jiménez, 2014)

Dentro de los otros tipos destacados se encuentra SEMMA, como “metodología enfocada especialmente en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando” (J. Mi. Moine et al., 2011), esta solo tiene en cuenta la muestra más representativa del problema.

De acuerdo a la comunidad KDnuggets (Data Mining Community's Top Resource) se realizó un comparativo de las metodologías usadas entre los años 2007 y 2014 (Figura 4) destacándose por su gran usabilidad y flexibilidad la metodología CRISP-DM, además de ser una de las metodologías que tiene en cuenta el análisis del problema, la planificación a futuro, actividades de planificación para el desarrollo de proyectos y actividades para el control o monitoreo del plan de trabajo (J. M. Moine et al., 2011).

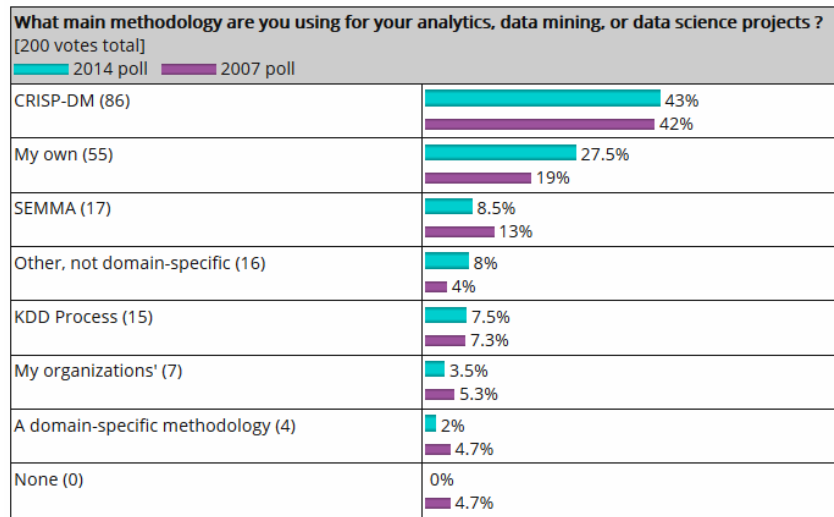


Figura 3. ¿Qué metodología usar para minería de datos?
Fuente: Piatetsky-Shapiro, G. (2014). KDnuggets

1.1.2. Investigaciones a nivel internacional

Grijalva, Freire, Real y Arellano (2018) realizaron un modelo predictivo de deserción según los factores que inciden en la eficiencia y deserción estudiantil de la Universidad Agraria de Ecuador, por lo tanto consideraron el uso de técnicas de minería de datos y la metodología KDD, para determinar patrones y relacionar atributos que inciden en la deserción académica. Para la predicción usaron arboles de decisión por medio del algoritmo c4.5 el cual:

Es un clasificador estadístico que utiliza la técnica de post-poda para mejorar la legibilidad del árbol. Se genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente, se construye con la estrategia de primero profundidad (depth-first), considerando todas las pruebas posibles al dividir los datos en conjuntos y se selecciona la prueba que haya generado la mayor ratio de ganancia. (Grijalva Arriaga et al., 2018)

Dentro de las técnicas de Clustering implementaron el algoritmo K-medias (k-means) con un total de tres cluster, los cuales tuvieron factores en relación con el ámbito académico (Promedio, número de materias aprobadas y el nivel de materias aprobadas) y aspectos socioeconómicos (edad y género) de los estudiantes. En conclusión, los algoritmos C4.5 y ID3 han permitido generar reglas que puedan predecir la deserción estudiantil.

Por otra parte, las investigaciones hechas por Vera, Romero y Ventura (2012) para predecir los factores que influyen en el fracaso o deserción de estudiantes de educación media, a través una muestra de 670 registros de estudiantes de los cuales 610 eran aprobados y 60 suspendidos, para el desarrollo implementaron una metodología propia conformada por 4 etapas:

- Recopilación de datos: recoger datos y determinar cuáles pueden ser de gran influencia. Debido al desequilibrio de los datos recogidos, emplearon SMOTE el cual “es un algoritmo de oversampling que genera instancias “sintéticas” o artificiales para equilibrar la muestra de datos basado en la regla del vecino más cercano” (Moreno et al., 2009).
- Pre-procesado: Uso de la técnica de ETL (Extraer, transformar y cargar) sobre los datos.
- Minería de datos: Dentro de esta etapa usaron algoritmos de clasificación y árboles de decisión, para determinar su exactitud y error, teniendo en cuenta 77 (Tabla 1) y 15 atributos (Tabla 2) respectivamente.

Tabla 2. Validación de datos con 77 atributos

Algoritmo	%Aciertos Aprobó	%Aciertos Suspendió	%Exactitud Total	Media Geométrica
JRip	97,7	78,3	96,0	87,5
NNge	98,5	73,3	96,3	85,0
OneR	98,9	41,7	93,7	64,2
Prism	99,5	25,0	93,1	49,9
Ridor	96,6	65,0	93,7	79,2
ADTree	99,7	76,7	97,6	87,4
J48	97,4	53,3	93,4	72,1
RandomTree	95,7	48,3	91,5	68,0
REPTree	98,0	56,7	94,3	74,5
SimpleCart	97,7	65,0	94,8	79,7

Fuente: (Vera, Romero & Ventura ,2012)

Tabla 3. Validación de datos con 15 atributos

Algoritmo	%Aciertos Aprobó	%Aciertos Suspendió	%Exactitud Total	Media Geométrica
JRip	97,0	81,7	95,7	89,0
NNge	98,0	76,7	96,1	86,7
OneR	98,9	41,7	93,7	64,2
Prism	99,2	44,2	94,7	66,2
Ridor	95,6	68,3	93,1	80,8
ADTree	99,2	78,3	97,3	88,1
J48	97,7	55,5	93,9	73,6
RandomTree	98,0	63,3	94,9	78,8
REPTree	97,9	60,0	94,5	76,6
SimpleCart	98,0	65,0	95,1	79,8

Fuente: (Vera, Romero & Ventura ,2012)

- Interpretación: Validez de la información y evaluación de los mejores resultados obtenidos por cada algoritmo.

Una de las conclusiones destacada por los investigadores es que “se ha mostrado que los algoritmos de clasificación pueden utilizarse con éxito para predecir el rendimiento académico de los estudiantes” (Vera, Romero & Ventura ,2012).

En estudios realizados para determinar indicadores académicos de una institución de educación superior, Camejo, Medina, Beltrán y Carrillo (2015) analizaron el comportamiento en cuanto a tiempo de respuesta, uso de CPU, uso de memoria y precisión, entre el algoritmo de regresión logística el cual es orientado a la predicción de una variable categórica en función de otras y árboles de decisión en donde se estable un nodo primario único y una representación esquemática de posibles alternativas. Para el análisis tuvieron en cuenta información como ingreso, matrícula, deserción, promoción, repitencia y eficiencia terminal de los estudiantes. Dentro de los resultados obtenidos evidenciaron que:

- El árbol de decisión obtuvo un tiempo de respuesta de 8.77 segundos menor que el de regresión lineal que presento 12.8 segundos.
- En cuanto al uso de la CPU el algoritmo de regresión logística un 0,58% mientras que el árbol de decisión uso un 0.65%
- En cuanto al uso de la RAM del árbol de decisión es de 133,18 MB; mientras que el de regresión lineal es de 133.68 MB.
- El algoritmo de árbol de decisión presenta una media de 0.75 de precisión; mientras que el de regresión lineal tiene una media de 0.74 de precisión.

En conclusión “se determinó como el algoritmo de mejor desempeño sobre datos académicos al algoritmo Árbol de Decisión” (Camejo et al., 2015)

Es así como “La predicción del rendimiento académico es una de las aplicaciones más populares de la EDM, debido a que el desempeño académico es uno de los principales índices de calidad académica de las instituciones educativas” (Rico Paez

& Sanchez Guzman, 2018), Es así como Rico y Sánchez (2018) se enfocaron en el análisis de información de 306 estudiantes de ingeniería del instituto Politécnico Nacional con el fin de implementar un modelo predictivo, el cual indique cuales son los factores que más influyen en el proceso de aprobación o reprobación del semestre, por consiguiente, usaron el algoritmo de Bayes Ingenuo, considerada una de las técnicas predictivas eficientes ya que “permite detectar problemas académicos con anticipación y tomar las decisiones más adecuadas”. Dentro de los análisis evidenciaron que la exactitud fue del 72.9% al incorporar los 12 mejores atributos (cursos reprobados, el tipo de curso, si tiene beca, el promedio actual, el promedio en media superior, entre otros), mientras que con un solo atributo obtuvieron un 65.27% de exactitud, tal como se evidencia en la gráfica al evaluar desde 1 a 20 atributos (Figura 5)

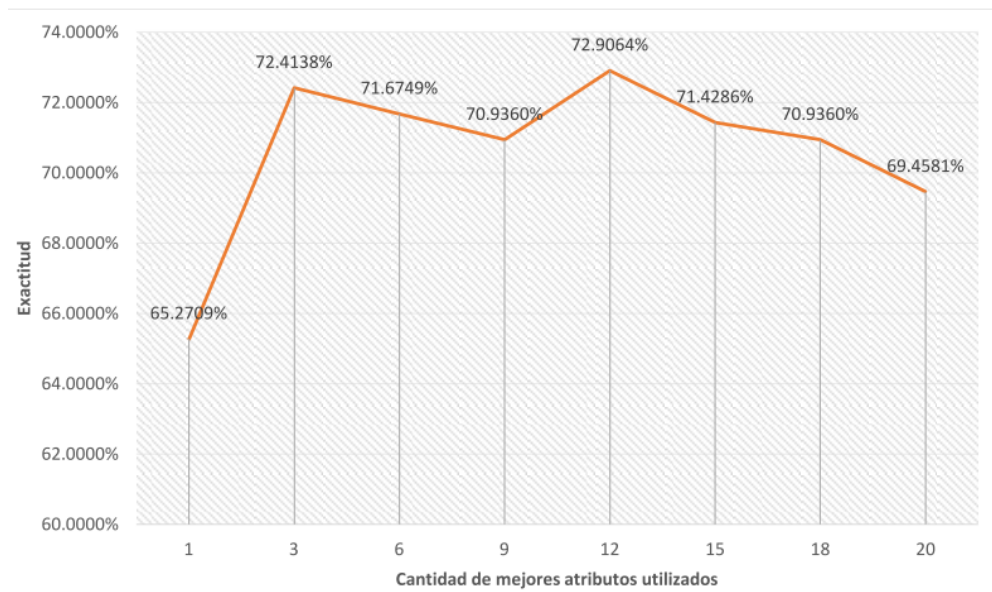


Figura 4. Exactitud de los resultados según los atributos usados
Fuente: (Oviedo Carrascal & Jiménez Giraldo, 2019).

1.1.3. Investigaciones a nivel nacional

En relación al proyecto de los factores asociados al desempeño académico en las pruebas saber 5° del 2014 al 2016, Gomez & Jaramillo (2017) indican que:

Según la Unesco, se evidencia una brecha de desigualdad en el rendimiento académico de los niños y niñas puesto que se ven reflejados los estereotipos de género que se marcan desde la familia y la sociedad, por ejemplo, la gran desventaja que se muestra en las niñas frente al aprendizaje de las matemáticas y a los niños en lectura.
(p.36)

Para encontrar los patrones asociados al bajo o alto desempeño, los investigadores utilizaron las técnicas descriptivas: reglas de asociación (Algoritmo Apriori), las cuales permiten identificar los factores que están incidiendo frecuentemente y su relación con la academia. Por otra parte, para agrupar los estudiantes según sus características emplearon el algoritmo K-means con el fin de obtener grupos homogéneos con semejanzas en los atributos identificados. Para la implementación de los algoritmos utilizaron la herramienta WEKA puesto que:

Contiene una colección de herramientas de visualización y algoritmos para el análisis de datos y el modelado predictivo, junto con interfaces gráficas de usuario para un fácil acceso a esta funcionalidad. Proporciona muchos algoritmos diferentes para el análisis de datos y el aprendizaje automático. Weka es de código abierto y está disponible gratuitamente. También es independiente de la plataforma.(Singhal & Jena, 2013)

En la predicción de los factores que afectan el rendimiento académico de los estudiantes en las pruebas Saber 11 de los años 2015 y 2016, Timarán, Caicedo y Hidalgo (2019) incorporaron la metodología CRISP-DM; dentro de la primera etapa (análisis del problema) indagaron sobre las posibles causas que inciden en el desempeño, posteriormente extrajeron 1.361.495 registros de las bases de datos del ICFES, y de esta manera realizaron su previa limpieza (eliminación de datos nulos) y transformación (selección de la información más relevante), para la aplicación de las técnicas de minerías de datos se tuvieron en cuenta del total 1.061.680 registros y 17 atributos, de los cuales se destacan el género, edad, estrato, SISBEN, ingreso familiar, educación de la madre y del padre; seguidamente procedieron a la construcción de los diferentes arboles de decisión de acuerdo al algoritmo j48 de la herramienta Weka, a su vez los investigadores usaron matrices de confusión con el propósito de determinar los aciertos y errores de clasificación. En conclusión, los resultados obtenidos indicaron que la creación de modelos interpretables de árboles de decisión son consistentes a la realidad y al respaldo teórico (Ricardo Timarán-Pereira et al., 2019).

“Se puede encontrar que las técnicas más comunes en EDM son árboles de decisión, redes bayesianas, regresión, correlación y análisis por clústeres, siendo el clustering la técnica más usada.” (Oviedo Carrascal & Jiménez Giraldo, 2019), ya que describe mejor los datos actuales. Para el análisis del desempeño de estudiantes de ingeniería en Antioquia en las pruebas Saber Pro Oviedo, Carrascal y Jiménez (2019) usaron la metodología CRISP-DM e implementaron los siguientes modelos:

1) **Clustering** en este proceso organizaron los clusters de acuerdo a los diferentes perfiles de estudiantes, por tanto, incorporaron al algoritmo k-vecinos el cual obtuvo un 81,12% de clasificación correcta (Figura 6)

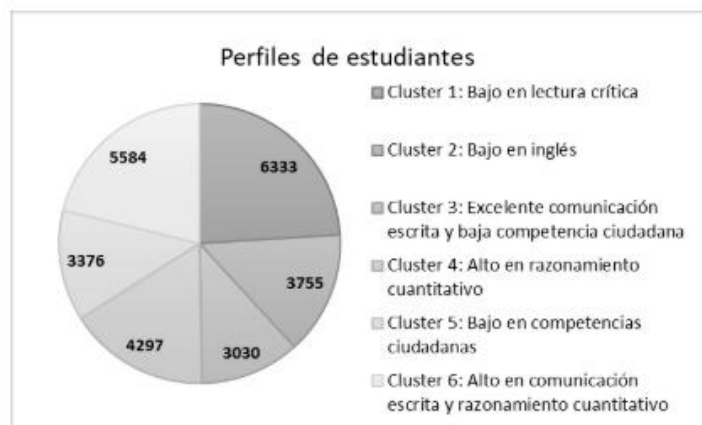


Figura 5. Perfiles de estudiantes
Fuente: (Oviedo Carrascal & Jiménez Giraldo, 2019).

2) **Selección de los factores que más influyen en el desempeño** para el modelo tuvieron en cuenta: inicialmente el análisis de correlación de Pearson, seguidamente el análisis de componentes principales para identificar que variables dependen de los resultados de cada módulo, de esta manera se identificaron 26 variables las cuales permitieron la predicción en un 81%.

3) **Predicción del desempeño en las pruebas a partir de las variables seleccionadas**; incorporaron arboles de decisión con el algoritmo C5.0, teniendo como referencia los cluster y las variables halladas.

En la siguiente tabla se observan otras investigaciones relacionadas al análisis del desempeño académico.

Tabla 4. Estudios relacionados con el rendimiento académico

País	Investigación	Técnicas
Argentina	Perfiles de Rendimiento Académico: Un Modelo basado en Minería de datos (D. Martínez, M. Karanik, M. Giovannini, 2015)	Arboles de Decisión y Clustering
Perú	Predicción del rendimiento académico aplicando técnicas de minería de datos, evaluando diversos algoritmos (Menacho Chiok, 2017)	Regresión Logística, Arboles de Decisión(J48), Redes Neuronales, Red Naive de Bayes
México	Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN (Rico Páez & Sánchez Guzmán, 2018)	Red Naive de Bayes
Colombia	La predicción del desempeño académico de los estudiantes de ingeniería usando los resultados de la prueba icfes y datos demográficos (Merchán Rubiano et al., 2017)	Arboles de decisión: J48 y Random Forest
Colombia	Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia	Red neuronal Perceptron Multicapa y Algoritmos de correlación

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

1.2. Línea de Investigación

El presente proyecto “Aplicativo web para la construcción de indicadores de brecha de género en la educación básica primaria y secundaria del país”, se realiza bajo la línea de investigación de Software, sistemas emergentes y nuevas tecnologías; la cual busca desarrollar proyectos investigativos que den soluciones tecnológicas a problemas y necesidades del entorno social, a partir del conocimiento y la innovación. Por tanto, se desarrollan productos tecnológicos que cumplan con algún objetivo de desarrollo sostenible.

1.3. Planteamiento del problema y pregunta de investigación

Las brechas de género han sido de gran importancia para la ONU, pues actualmente en el mundo existen estereotipos que impiden el crecimiento y desarrollo del país, es así como los Objetivos de Desarrollo Sostenible se centran en el mejoramiento de aspectos sociales, económicos, políticos y ambientales de las personas, por ello dentro del Objetivo 4 “Educación de Calidad” se busca para el año 2030:

- Asegurar el acceso igualitario de todos los hombres y las mujeres a una formación técnica, profesional y superior de calidad, incluida la enseñanza universitaria
- Eliminar las disparidades de género en la educación y asegurar el acceso igualitario a todos los niveles de la enseñanza y la formación profesional para las personas vulnerables, incluidas las personas con discapacidad, los pueblos indígenas y los niños en situaciones de vulnerabilidad.
- Asegurar que todas las niñas y todos los niños terminen la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados de aprendizaje pertinentes y efectivos

En relación al objetivo 5 “Igualdad de Género” el cual propone:

- Poner fin a todas las formas de discriminación contra todas las mujeres y las niñas en todo el mundo
- Asegurar la participación plena y efectiva de las mujeres y la igualdad de oportunidades de liderazgo a todos los niveles decisorios en la vida política, económica y pública
- Aprobar y fortalecer políticas acertadas y leyes aplicables para promover la igualdad de género y el empoderamiento de todas las mujeres y las niñas a todos los niveles.

En Colombia se evidencia que la implementación de algunos de los Objetivos de Desarrollo Sostenibles están en espera o proceso, tales como educación e igualdad de género, pues así lo evidencia el informe de niveles de paridad entre hombres y mujeres del Foro Económico Mundial (2018), en donde Colombia ocupa la posición 40 entre 149 países, indicando que uno de los factores que influye en el resultado son los roles que desempeñan las mujeres dentro de las organizaciones y la relación con su respectiva remuneración. En parte esta problemática inicia desde el proceso educativo ya que para London et al. (2011):

La institución educativa juega un papel fundamental en el pronóstico del desarrollo profesional de las mujeres dentro de estos recintos, pues, desde su instancia como tal, es ella la que influye en la magnitud de la ausencia o presencia de roles femeninos en puestos directivos y de toma de decisiones, en la determinación de salarios y prestaciones, así como de requisitos para su promoción y ascenso laboral en términos de carrera profesional y de desarrollo profesional dentro de los recintos educativos; por lo que el reto es abogar, desde todos los medios posibles que rijan ese parámetro, por alcanzar una

equidad de género. Tomado de la investigación de: (Ojeda García, 2014, p.72)

Para determinar el comportamiento académico en la educación básica primaria y secundaria, el ICFES durante el año 2009 diseñó las pruebas 3°, 5° y 9°, con el fin de determinar los avances que tiene la educación colombiana. Estas pruebas evalúan las competencias en matemáticas, lenguaje, ciencias naturales y competencias ciudadanas para 5° y 9°, se desarrollaron en el marco de los Estándares Básicos de Competencias teniendo como referencia lo que deben saber y saber hacer los estudiantes. Dado que los estudios en torno a los indicadores de brecha de género en Colombia son muy pocos y que los datos suministrados por el ICFES, no aportan conocimiento para la toma de decisiones, se ve necesario el uso de metodologías ingenieriles para explorar más a fondo el comportamiento de los datos, ya que algunos análisis estadísticos solo se encargan de comparar grupos y el resultado no es profundo.

¿Cómo afecta el resultado de las pruebas saber de la educación básica primaria y secundaria en las brechas de género del país?

1.4. Objetivo general y objetivos específicos

Desarrollar un aplicativo web usando técnicas de minería de datos, para la construcción de indicadores de brechas de género en el país, de acuerdo, a los datos académicos, socioeconómicos y familiares de los resultados de las pruebas Saber 3°, 5° y 9°.

- Realizar un previo análisis de la información más importante reposada en las bases de datos del ICFES.

- Realizar extracción, limpieza y transformación de los datos para su uso dentro de la minería de datos.
- Implementar un modelo descriptivo por medio de herramientas especializadas, el cual permita generar indicadores de las brechas de género de la educación primaria y media del país.
- Realizar escenarios de prueba para el análisis de los modelos desarrollados, para corroborar el margen de exactitud y error presentados.
- Implementar un modelo predictivo con el uso de técnicas de minería de datos el cual incorpore los resultados académicos de todos los niveles de educación.

1.5. Alcance e impacto del proyecto

En el marco de la investigación de brechas de género en el país, se destacan los Objetivos de Desarrollo Sostenible planteados por la ONU tales como:

- **Objetivo 5** “Lograr la igualdad entre los géneros y empoderar a todas las mujeres y las niñas”, teniendo en cuenta la participación política, la reducción de la tasa desempleo, la eliminación de la violencia, la igualdad de oportunidades y por supuesto, el aumento en los ingresos salariales de las mujeres.
- **Objetivo 4** “Educación de calidad”. Dentro de este objetivo se plantea como meta principal para el 2030, eliminar las disparidades de género en la educación y asegurar el acceso igualitario a todos los niveles de enseñanza y formación profesional.

Por consiguiente, con ayuda del portal de datos abiertos del ICFES se desarrollará un modelo descriptivo que mida la brecha de género e identifique los factores asociados a la problemática, posteriormente incorporar la información obtenida en

un modelo predictivo, el cual ayude en la toma de decisiones para la mejora en la calidad y la equidad de la educación colombiana desde los centros educativos hasta la sociedad en general.

Siendo la mujer un ente investigador que genera aportes significativos en el desarrollo e innovación de proyectos científicos y tecnológicos, los cuales aportan al crecimiento económico del país; por ello, no solamente se pretende contribuir en la mejora de la educación sino desde el ámbito educativo generar ideas que ayuden al aumento de participación laboral y competitividad de género igualitaria.

1.6. Metodología

El proyecto sigue los lineamientos de la metodología cuantitativa, la cual parte de la recolección de datos que se analizan mediante procesos estadísticos, y de esta manera se permitan probar hipótesis, establecer patrones y probar teorías, esta metodología es caracterizada por emplear procesos cuidadosos, sistemáticos y empíricos en su esfuerzo por generar conocimiento (Hernández-Sampieri et al., 2006).

En cuanto a los métodos de gestión de proyecto se implementara la metodología de desarrollo ágil SCRUM, para ello se incorporaran los principios según las necesidades del cliente y el conocimiento tácito de las personas, lo cual conlleve al desarrollo incremental de iteraciones de acuerdo a la organización y colaboración del equipo, para evaluar el progreso se realizaran Sprints con el fin de mostrar al cliente los avances del producto. Se dará por terminado el proceso cuando el cliente lo decida (Palacios, 2015).

Dentro de la metodología propuesta se incorporara el uso de las técnicas descriptivas y predictivas de la minería de datos, tales como clustering. Para ello

se tendrá en cuenta el modelo CRISP-DM “considerada la metodología más usada en el desarrollo de proyectos en los ambientes académico e industrial” (Timaran-Pereira et al., 2017), partiendo en primera instancia con el análisis del negocio el cual se enfoca en determinar los objetivos del negocio según el contexto y la situación actual, en segunda instancia la comprensión de los datos, en la cual se evalúa los atributos y las variables más importantes, en tercera instancia la preparación de los datos, aquí se hace una selección y limpieza de los datos, con el fin de obtener datos más consistentes y equilibrados para el desarrollo, en cuarta instancia el modelado, en esta fase se incorporaran técnicas de minería de datos acorde a lo datos seleccionados anteriormente, en quinta instancia la evaluación en esta fase se valida el modelo desarrollado y por último la explotación en el cual se incorpora el modelo dentro de una aplicación para la demostración al usuario final.

1.7. Marcos de Referencia

1.7.1. Marco Teórico

En esta sección se describen conceptos importantes y explicaciones teóricas relacionadas con el desarrollo del proyecto tales como:

1.7.1.1. Brechas de género

Las brechas de género se manifiestan desde el ámbito académico; las niñas se van alejando progresivamente de las áreas del STEM (Ciencia, tecnología, ingeniería y matemáticas) debido a múltiples causas como pensamientos familiares, estereotipos, expectativas y falta de modelos a seguir. De acuerdo a la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), muestra que en gran parte de los países de América Latina y el Caribe, las niñas desde sexto grado obtienen bajos resultados en matemáticas que los niños, es por

tal motivo que a la hora de elegir carreras profesionales existe una menor inclinación en áreas del CTI(Ciencia, tecnología e innovación), además de ello su participación en posgrados en disciplinas del STEM tiende a disminuir a nivel mundial (López et al., 2018).

De manera que en varios países para muchas mujeres la integración entre la vida profesional y familiar es más difícil y por tanto, las brechas salariales persisten en la mayoría de las profesiones científicas y en la presencia en cargos gerenciales es muy baja con respecto a los hombres, numerosos estudios indican que la baja participación femenina tiene un costo económico importante para los países ya que es un potencial subutilizado. Por ende organismos internacionales identifican 16 indicadores que cubren la totalidad de la problemática relacionada a la brecha de género las cuales se organizan por cuatro áreas temáticas: educación superior en disciplinas del STEM, carreras en ciencia y tecnología, investigación científica e innovación y emprendimiento innovador (López et al., 2018).



Figura 6. Indicadores de brechas de género
Fuente: (López et al., 2018).

1.7.1.2. Minería de datos

Es un proceso en el cual se extrae conocimiento útil y comprensible, mediante métodos y técnicas, a partir de grandes volúmenes de datos, con el objeto de predecir de forma autorizada tendencias y comportamientos o descubrir de forma automatizada modelos previamente desconocidos. Por otra parte, una definición más concreta indica que la minería de datos se centra en la búsqueda de patrones, relaciones, reglas y excepciones útiles las cuales son obtenidas mediante árboles o reglas de clasificación, técnicas de regresión, clusterizado, modelizado secuencial, dependencias, etc. (Rodríguez Suárez & Amador, 2009).

De acuerdo a Rodríguez y Amador (2009) las técnicas de minería de datos han sido usadas desde hace varios años para la obtención de patrones en los datos y para la extracción de información valiosa, por tanto ha sido usado para el clima, la medicina, la mercadotecnia, detención fraudes y análisis financieros. Dichas técnicas se dividen en dos grandes ramas:

1. Supervisados o predictivos: Se predice el valor de un atributo a partir de un conjunto de datos conocidos.
2. No supervisados o del descubrimiento del conocimiento: Se descubren y analizan patrones o tendencias en los datos actuales.

Así mismo los autores Rodríguez y Amador indican que las técnicas de recopilación de datos se usan herramientas como:

1. Almacenamiento de datos (Data Warehousing): Por medio de la organización de grandes volúmenes de datos se logra determinar o referenciar conjuntos de datos importantes para un proceso de toma de decisión dado.

2. Análisis exploratorio de datos (Exploratory Data Analysis): Determinar las relaciones entre las variables cuando no hay o no está definida la naturaleza de dichas relaciones. Están diseñadas para identificar patrones en conjunto de datos multivariantes.

1.7.1.3. Minería de datos educacional

Las técnicas de minería de datos son usadas para la determinar y comprender el comportamiento de los estudiantes. La MDE emerge como un paradigma orientado para la generalización de modelos, tareas, métodos y algoritmos para el análisis de datos que provienen de un contexto educativo asimismo tiene como función encontrar, analizar patrones que caractericen los comportamientos en base a sus logros, evaluaciones y el dominio de contenido de conocimiento que tienen los alumnos en los diversos mecanismos de aprendizaje-enseñanza que hoy en día son otorgados en las diversas instituciones con el objetivo de generar modelos educativos en los cuales se puedan fomentar nuevas herramientas que se puedan analizar según el nivel participativo de los estudiantes sobre los sistemas aprendizaje-enseñanza. Para analizar dichos comportamientos se debe tener en cuenta que no solo existe información de manera presencial, ya que en el ámbito educativo existen diferentes modalidades de aprendizaje tales como: presencial, que es cuando los estudiantes acuden a las escuelas y los docentes exponen sus temas; la segunda es la mixta, donde los estudiantes en algunos periodos asisten presencialmente y en otros desarrollan actividades virtuales; por último se habla de la modalidad virtual o en línea donde el estudiante y el docente interactúan solo por internet. Es por tal motivo que la educación enfrenta un gran reto al querer pronosticar las trayectorias que generan los estudiantes en la iteración que estos tienen con los sistemas educativos de aprendizaje y de cómo es el flujo de información (Ballesteros et al., 2013).

1.7.1.4. CRISP-DM

Según Moine, Gordillo y Haedo (2011) esta metodología fue creada por las empresas de SPSS, NCR y Daimler Chrysler en el 2000, es una de las metodologías más usadas en el desarrollo de proyectos de minería de datos, adicionalmente establece un conjunto de tareas y actividades para cada fase del proyecto pero no especifica cómo tratarlas. Se estructura en seis etapas:

1. Comprensión del negocio: Evaluación del problema a abordar y el contexto organizacional
2. Comprensión de los datos: Análisis de las variables y los datos, para determinar cuáles serán usadas.
3. Preparación de los datos: Limpieza, pre-procesamiento, reducción y proyección de los datos
4. Modelado: Selección de técnicas, modelos u algoritmos de minería de datos.
5. Evaluación e implementación: Interpretación, refinamiento del modelo, difusión del nuevo conocimiento dentro de la organización, planificación para el control futuro y un análisis de cierre del proyecto.

1.7.1.5. Clustering o Agrupamiento

A diferencia de la clasificación y la predicción que utilizan modelos supervisados analizan los registros con etiquetas o categorías los objetos de los datos, de tal manera que las técnicas de agrupamiento no se interesan por modelar un conjunto de relaciones y respuestas que pertenezcan a cada valor asociado. Esta técnica utiliza el modelado no-supervisado donde analiza los objetos sin consultar con una clase conocida, en pocas palabras, las etiquetas de la clase no se encuentran en

los datos de entrenamiento por el hecho de que el objeto es desconocido (Ballesteros et al., 2013).

Esta técnica es utilizada para la generalización de etiquetas esto es la búsqueda y construcción de grupos formados de tal forma que los patrones de un grupo tienen una alta similitud en función de la medida de asociación, donde cada conjunto construido puede ser interpretado como una clase de objetos de los cuales se pueden establecer reglas (Ballesteros et al., 2013).

En términos más simples Clustering es la división o partición de un conjunto de observaciones con el objetivo principal de asignar cada uno de los N objetos de datos a uno de los K posibles grupos disjuntos mediante una medida de similitud, cuyo objetivo es descubrir y modelar los grupos en que los elementos de datos se agruparan, es decir se encarga de agrupar los datos en categorías más comunes (Ballesteros et al., 2013).

1.7.1.6. Árboles de decisión

Es un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Son muy parecidas a los sistemas de predicción basadas en reglas, los cuales sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva para la solución de un problema (Barrientos et al., 2009).

El conocimiento obtenido durante el proceso de aprendizaje inductivo se presenta mediante de un árbol, el cual se representa por un conjunto de nodos, hojas y ramas. Cada nodo principal es el atributo a partir del cual se inicia el proceso de clasificación; los nodos internos corresponden a las preguntas relacionadas con el atributo principal del problema; cada una de las respuestas a dichos cuestionamientos se representa por un nodo hijo. Las ramas que salen de cada

nodo se encuentran etiquetadas con los posibles valores del atributo. Los nodos finales corresponden a una decisión, la cual coincide con una de las variables clase del problema a resolver (Barrientos et al., 2009).

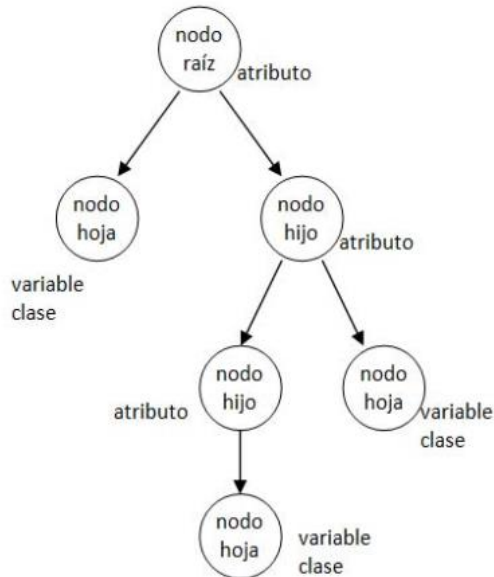


Figura 7. Esquema de un árbol de decisión
Fuente: (Barrientos et al., 2009).

1.7.1.7. WEKA

Es de libre distribución (Licencias GPL), es una herramienta destacada por la gran cantidad de algoritmos que contiene, así mismo es destacado por su eficiencia a la hora de implementarlos, fue desarrollada por investigadores de la Universidad de Waikato, adicionalmente proporciona una gran cantidad de herramientas para la realización y el desarrollo de modelos a partir de la minería de datos, también permite la programación en JAVA de algoritmos más sofisticados para el análisis de datos, todo esto se reúne en una interfaz gráfica, la cual permite acceder fácilmente a sus funcionalidades. En ella se implementan las técnicas de clasificación, asociación, agrupamiento, y predicción existentes en la actualidad. Su sistema operativo es multiplataforma (Rodríguez Suárez & Amador, 2009).

1.7.1.8. JAVA

De acuerdo a Garrido (2015) Java es un lenguaje de programación de alto nivel orientado a objetos de reciente creación, el cual fue desarrollado por Sun Mycosystem a principios de los años 90, y se caracteriza por:

- Ser un lenguaje familiar, es decir que fue desarrollado desde cero como un nuevo lenguaje pero es muy similar a C o C++, por tanto facilita la migración de aquellos desarrolladores ya familiarizados con dichos lenguajes
- Ser un lenguaje sencillo, el conocer otros lenguajes permite un mejor aprendizaje, adicionalmente sirve para el desarrollo en distintos dispositivos.
- Multiplataforma: Es decir se escribe y compila una sola vez en una plataforma y se ejecuta en otra sin tener que modificar el código fuente. Java es un lenguaje interpretado, aunque también compilado para mejorar el rendimiento de la interpretación.
- Java es un lenguaje de alto rendimiento
- Robusto: Java se encarga de comprobar la sintaxis algo que otros lenguajes no logran hacer. Durante la ejecución se vuelve a realizar otras comprobaciones, la cual arroja resultados inesperados y de esta manera le permite al desarrollador realizar mejoras.

1.7.1.9. PostgreSQL

Es un potente motor de base de datos relacionales reconocido por su fiabilidad, integridad de datos y correcto desempeño, como así también por su alta portabilidad a los principales sistemas operativos Linux, Unix y Windows. Su código fuente está disponible bajo licencia de código abierto por lo que es posible su uso, modificación y distribución (Sarría, n.d.).

1.7.2. Marco Legal

Este proyecto se desarrollara en términos de software libre, es decir que no requiere de licencias privadas, por tal motivo se le permite a los usuarios realizar estudios y modificaciones las cuales permitan mejorar sus capacidades y adaptarlo a necesidades particulares. De acuerdo a Rodríguez (1992) el autor otorga las siguientes libertades a los usuarios:

- La libertad de ejecutar el programa para cualquier propósito.
- La libertad de estudiar la manera en que el programa opera y adaptarlo a sus necesidades particulares.
- La libertad para redistribuir copias del programa (incluido su código fuente) a quien desee.
- La libertad de mejorar el programa y distribuir sus mejoras al público bajo las mismas condiciones del programa original.

A continuación se describen aspectos legales y normativos que influyen en el proyecto en cuanto a minería de datos, derechos de autor y protección de obras.

La constitución política de Colombia en el artículo 61, establece que “El estado protegerá la propiedad intelectual por el tiempo y mediante las formalidades que establezca la ley”.

1.7.2.1. Ley 23 de 1982

Artículo 1:

Autores de obras literarias, científicas y artísticas gozarán de protección para sus obras en la forma prescrita por la presente ley y, en cuanto fuere compatible con ella, por el derecho común. También protege esta ley a los

intérpretes o ejecutantes, a los productores de fonogramas y a los organismos de radiodifusión, en sus derechos conexos a los del autor.

Artículo 2:

Los derechos de autor recaen sobre las obras científicas, literarias y artísticas las cuales se comprenden todas las creaciones del espíritu en el campo científico, literario y artístico, cualquiera que sea el modo o forma de expresión y cualquiera que sea su destinación, tales como: los libros, folletos y otros escritos; las conferencias, alocuciones, sermones y otras obras de la misma naturaleza; las obras dramáticas o dramático musicales; las obras coreográficas y las pantomimas; las composiciones musicales con letra o sin ella; las obras cinematográficas, a las cuales se asimilan las obras expresadas por procedimiento análogo a la cinematografía, inclusive los video-gramas; las obras de dibujo, pintura, arquitectura, escultura, grabado, litografía; las obras fotográficas a las cuales se asimilan las expresadas por procedimiento análogo a la fotografía; las obras de arte aplicadas; las ilustraciones, mapas, planos, croquis y obras plásticas relativas a la geografía, a la topografía, a la arquitectura o a las ciencias, y, en fin, toda producción del dominio científico, literario o artístico que pueda reproducirse, o definirse por cualquier forma de impresión o de reproducción, por fonografía, radiotelefonía o cualquier otro medio conocido o por conocer.

Artículo 30:

El autor tendrá sobre su obra un derecho perpetuo, inalienable, e irrenunciable para:

- a) Reivindicar en todo tiempo la paternidad de su obra y, en especial, para que se indique su nombre o seudónimo cuando se realice cualquiera de los actos mencionados en el artículo 12 de esta ley.

- b) A oponerse a toda deformación, mutilación u otra modificación de la obra, cuando tales actos puedan causar o acusen perjuicio a su honor o a su reputación, o la obra se demerite, y a pedir reparación por éstos.
- c) A conservar su obra inédita o anónima hasta su fallecimiento, o después de él cuando así lo ordenase por disposición testamentaria.
- d) A modificarla, antes o después de su publicación, y
- e) A retirarla de la circulación o suspender cualquier forma de utilización aunque ella hubiese sido previamente autorizada

1.7.2.2. El Decreto 1360 de 1989

“Por el cual se reglamenta la inscripción del soporte lógico (software en el Registro Nacional del Derecho de Autor)”, incorporó el concepto del software en la normativa colombiana. Este decreto define al software “como una creación propia del dominio literario” comprendiendo el programa de computador (u ordenador), su descripción y el material auxiliar. Sin embargo, el artículo 7 señala que “la protección otorgada al soporte lógico, no excluye otras formas de protección”, es decir que otros componentes del software están protegidos.

1.7.2.3. Decisión 351 de 1993 Acuerdo de Cartagena

De acuerdo al Régimen Común sobre Derechos de Autor y Derechos Conexos. En su capítulo 1 indica que se protege el programa de ordenador (software) la cual define como “Expresión de un conjunto de instrucciones mediante palabras, códigos, planes o en cualquier otra forma que, al ser incorporada en un dispositivo de lectura automatizada, es capaz de hacer que un ordenador (un aparato electrónico o similar capaz de elaborar informaciones) ejecute determinada tarea u

obtenga determinado resultado. El programa de ordenador comprende también la documentación técnica y los manuales de uso”.

1.7.2.4. Ley 565 de 2000

Por medio de la cual se aprueba el Tratado de la OMPI sobre Derechos de Autor, adoptado en Ginebra el 20 de diciembre de 1996. Artículo 4: “Los programas de ordenador están protegidos como obras literarias en el marco de lo dispuesto en el artículo 2 del Convenio de Berna. Dicha protección se aplica a los programas de ordenador, cualquiera que sea su modo o forma de expresión.”

1.7.2.5. Ley 1712 de 2014

Se podrá hacer uso, aprovechamiento y/o transformación de forma libre y sin restricciones, para hacer aplicaciones por parte de terceros y contenidos de su propia creación

2. AUTOMATIZACIÓN DE DESCARGA DE PRUEBAS

Para la descarga de los archivos CSV del FTP del ICFES y la carga de los mismos a la base de datos, se realizaron los siguientes procedimientos:

2.1. Implementación de WinSCP

Por medio del cliente SFTP gráfico de WinSCP y con tareas programadas se buscó generar automáticamente la descarga de resultados, para ello se tuvieron en cuenta los siguientes pasos:

- Instalar la aplicación WinSCP
- Crear dos carpetas en el disco C, una para guardar los scripts creados y otra para guardar los archivos descargados del FTP.
- Posteriormente se crea un archivo denominado “sync_winscp.txt”, el cual se guardara en la carpeta de los scripts y contendrá información relacionada con la ruta, el usuario y la contraseña de acceso al FTP, adicionalmente contendrá la ruta de la carpeta donde se descargarán los archivos.

```

sync_winscp: Bloc de notas
Archivo Edición Formato Ver Ayuda

option batch abort
option confirm off
open ftp://ftpdbicfes:dataices$%@ftp.icfes.gov.co
synchronize remote c:\Sincronizacion_FTP/
exit

```

Figura 8. SCRIPT de conexión al FTP
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Una vez creado el script se procede a crear un archivo bat denominado “sync_winscp.bat”, el cual contendrá la ruta de acceso del script creado anteriormente, así mismo este archivo será almacenado en la misma carpeta del script.

```

sync_winscp: Bloc de notas
Archivo Edición Formato Ver Ayuda

@echo off
"C:\Program Files (x86)\WinSCP\winscp.com" /script=sync_winscp.txt

```

Figura 9. Archivo BAT para localizar script
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- A continuación se debe ingresar al programador de tareas, con el fin de ejecutar en una hora del día el archivo “bat” creado anteriormente.
- Una vez ingresado se debe seleccionar la opción de “Crear tarea”.
- Luego se asigna un nombre a la tarea y se realiza la siguiente configuración.

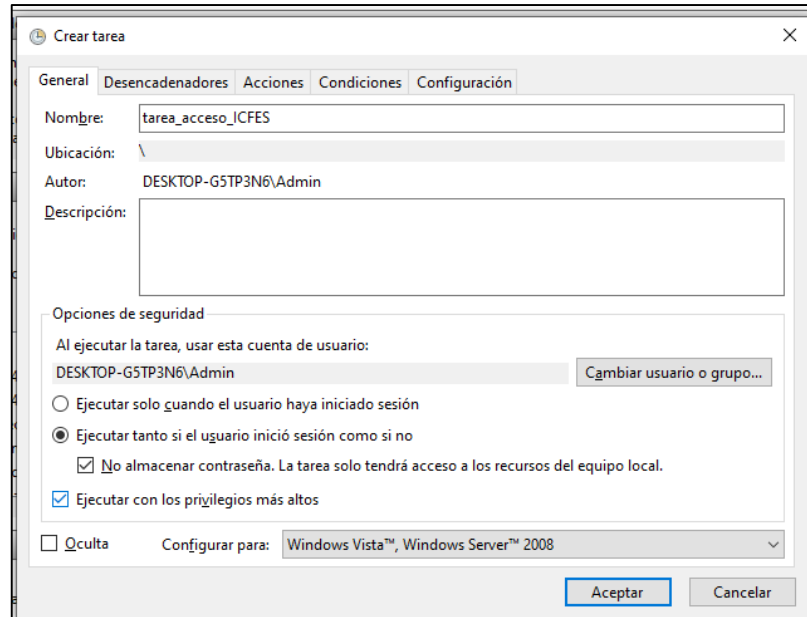


Figura 10. Configuración de tarea programada
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Posteriormente se seleccionara la pestaña de “Desencadenadores” y la opción “nuevo”, allí se programara el tiempo de ejecución de la tarea.

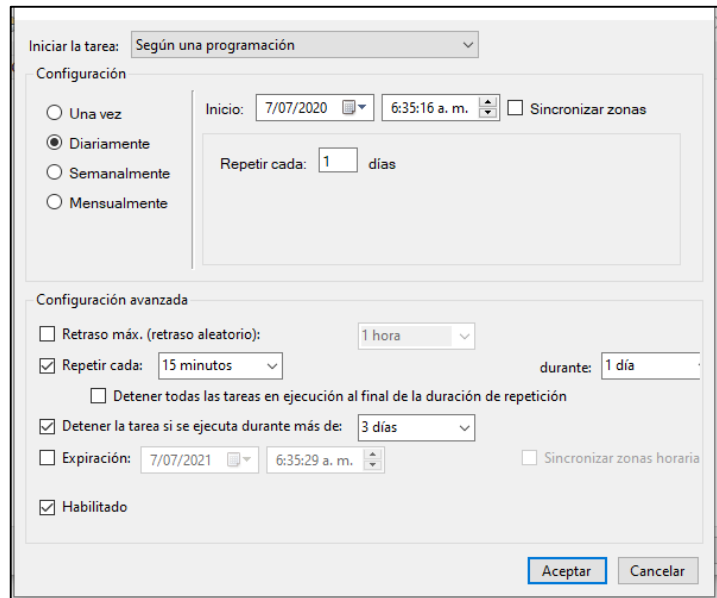


Figura 11. Configuración de tiempos de la tarea programada
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Luego de crear el desencadenador, se configurará la ruta del “bat” que se quiere ejecutar

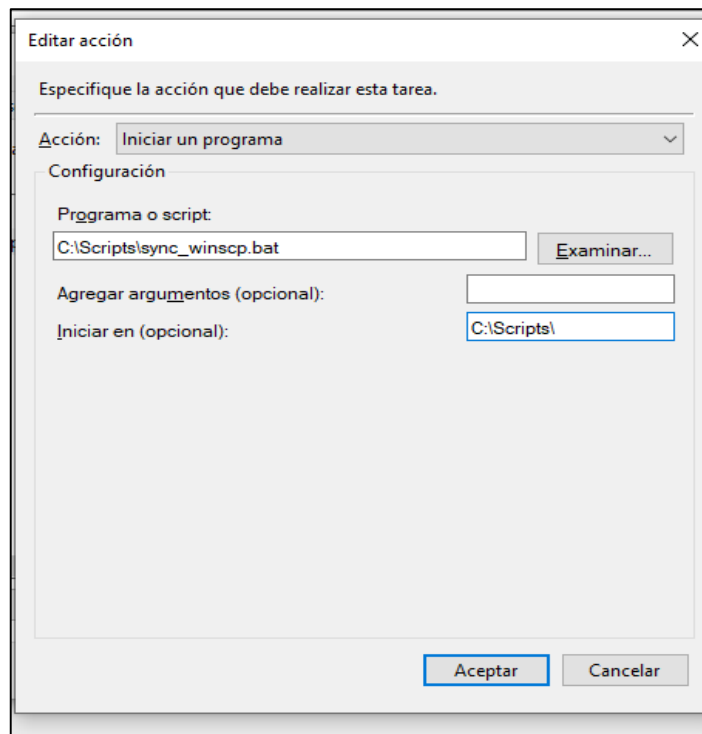


Figura 12. Configuración de la ruta del “bat”
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Por último, se verifica la tarea creada y se ejecuta para comprobar su funcionamiento.

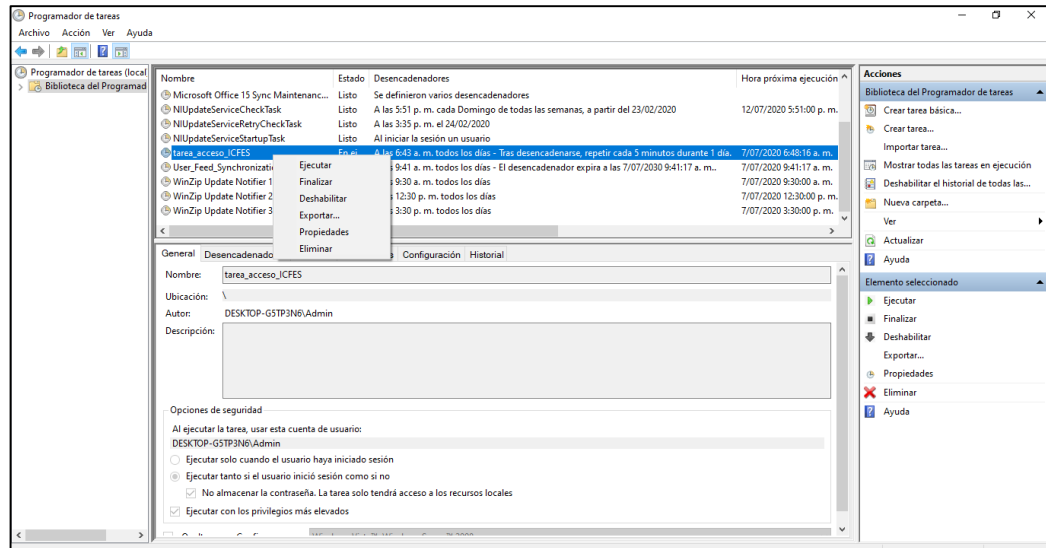


Figura 13. Ejecución de tarea programada
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

2.2. Configuración de AutoFTP Manager

AutoFTP Manager es una herramienta de transferencia de archivos administrada que facilita la programación y descarga de archivos de un FTP. Para instalarla y configurarla se deben seguir los siguientes pasos:

- Instalar Auto FTP Manager en el equipo
- Crear un nuevo perfil de conexión, para este paso se debe tener el usuario, contraseña y ruta de acceso al FTP del ICFES, así se configurara la herramienta:

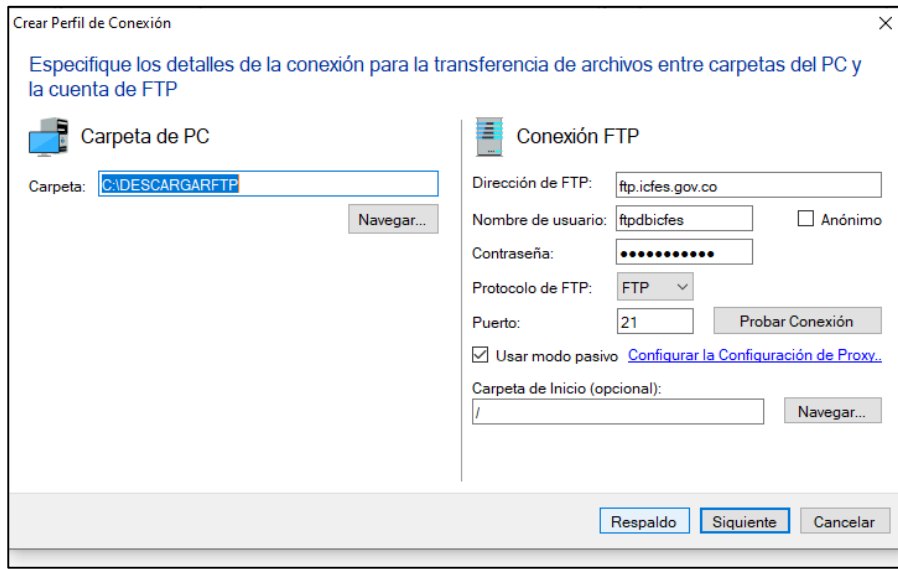


Figura 14. Configuración de perfil de conexión
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Una vez creado el perfil se añadirán reglas de transferencia automatizada

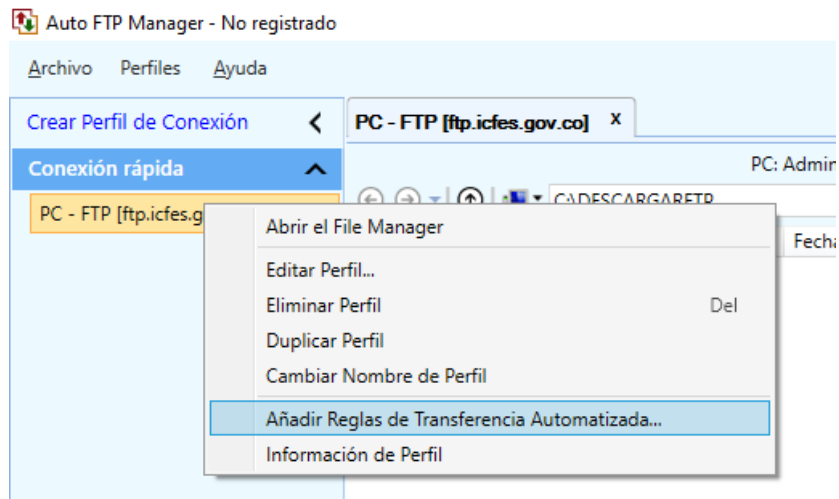


Figura 15. Configuración de perfil de conexión
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Posteriormente, se configura el tipo de transferencia, el tiempo y las carpetas que desea sincronizar.

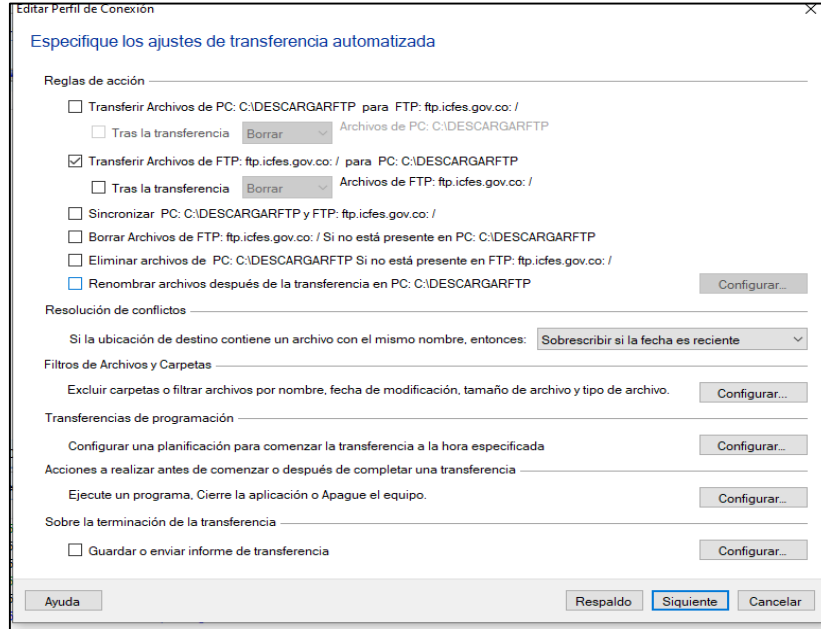


Figura 16. Configuración de ajustes de transferencia
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

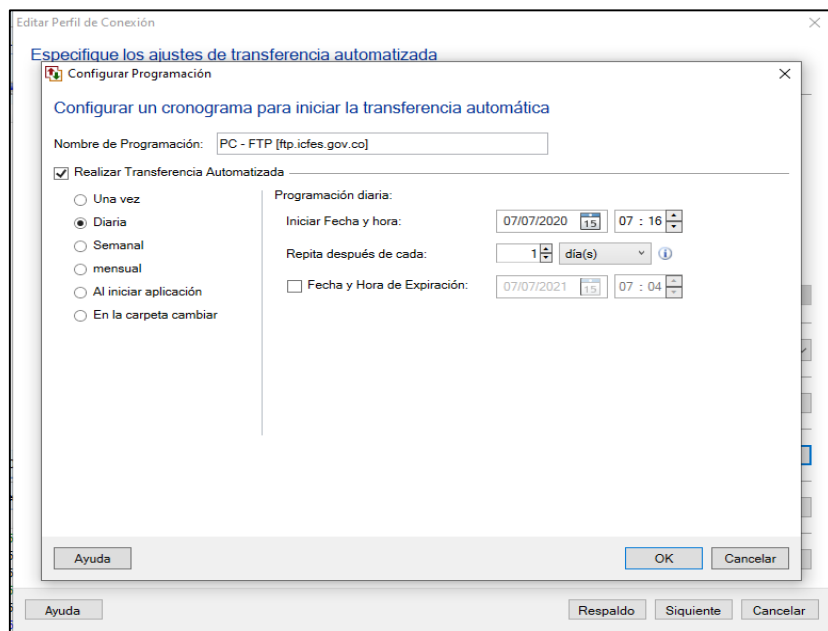


Figura 17. Configuración de programación de tiempos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Una vez configurado se procede a ejecutar la tarea programada de acuerdo a la configuración establecida.

Para el proceso de automatización de descarga de las pruebas saber con las herramientas Auto FTP Manager y WinSCP, se presentaron fallas debido a la inconsistencia del FTP del ICFES, ya que la conexión era intermitente y presentaba fallas a la hora de extraer los documentos. No obstante, para el año 2020 el ICFES traslado la información del repositorio FTP a un OneDrive.

3. SELECCIÓN DE INDICADORES

3.1. Configuración de WEKA

Para la selección de los indicadores se hizo uso del aplicativo Weka desde allí se realizó el análisis de cada una de las variables cargadas en el archivo CSV, inicialmente se realizó la siguiente configuración:

- Instalar WEKA
- Descargar el controlador de PostgreSQL, versión “42.2.9”
- Adicionar a la carpeta donde quedo instalado el WEKA
- Buscar el archivo “DatabaseUtils.props” y configurarlo de la siguiente manera

```
# url: http://www.mysql.com/
# jdbc: http://www.mysql.com/products/connector/j/
# author: Fracpete (fracpete at waikato dot ac dot nz)
# version: $Revision: 15255 $

# JDBC driver (comma-separated list)
jdbcDriver=org.postgresql.Driver

# database URL
jdbcURL=jdbc:postgresql://localhost:5432/basededatos

# specific data types
string, getString() = 0; --> nominal
boolean, getBoolean() = 1; --> nominal
double, getDouble() = 2; --> numeric
byte, getByte() = 3; --> numeric
short, getByte() = 4; --> numeric
int, getInteger() = 5; --> numeric
long, getLong() = 6; --> numeric
float, getFloat() = 7; --> numeric
date, getDate() = 8; --> date
text, getString() = 9; --> string
time, getTime() = 10; --> date
timestamp, getTime() = 11; --> date

# other options
```

Figura 18. Configuración del archivo “DatabaseUtils.props”
Fuente: (Viracacha Viracacha & Barahona Rodríguez, 2020)

- Una vez configurado, abrir weka y seleccionar la opción “Explorer”

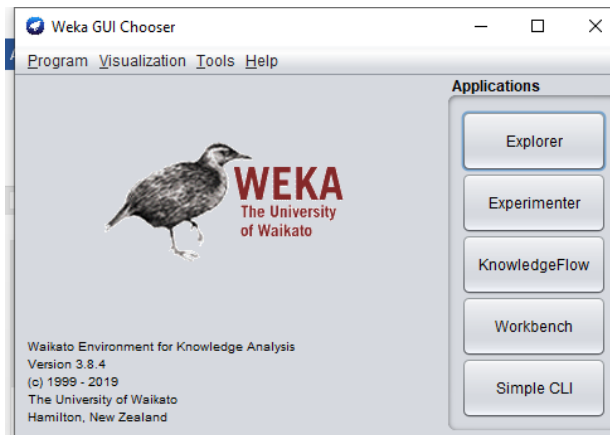


Figura 19. Interfaz inicial de weka
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- A continuación, seleccionar la opción “Open DB”, y la opción “set user and password”, para registrar la URL, el usuario y la contraseña de la base de datos:

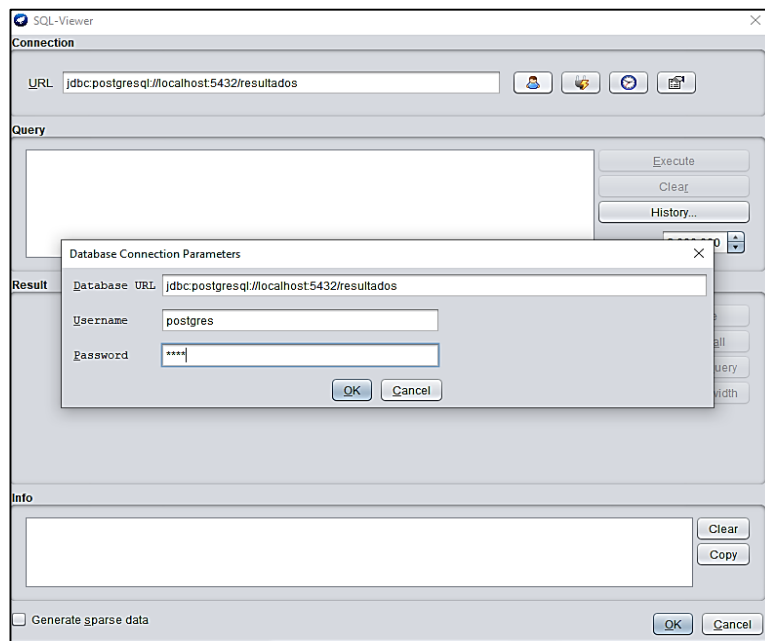


Figura 20. Configuración de usuario y contraseña
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Una vez establecida la conexión con la base de datos, se procede a registrar el query de consulta.

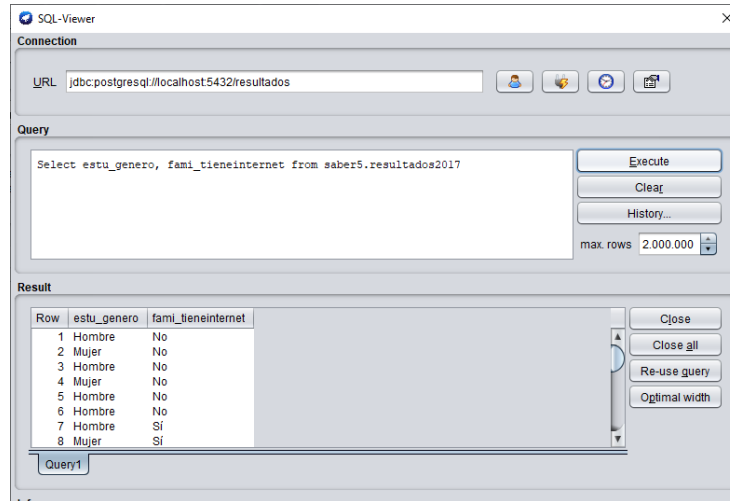


Figura 21. Query de consulta
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- A partir de los datos obtenidos se debe seleccionar el tipo de análisis a realizar. En este caso se seleccionó la metodología de “Cluster” y como algoritmo “SimpleKMeans”.

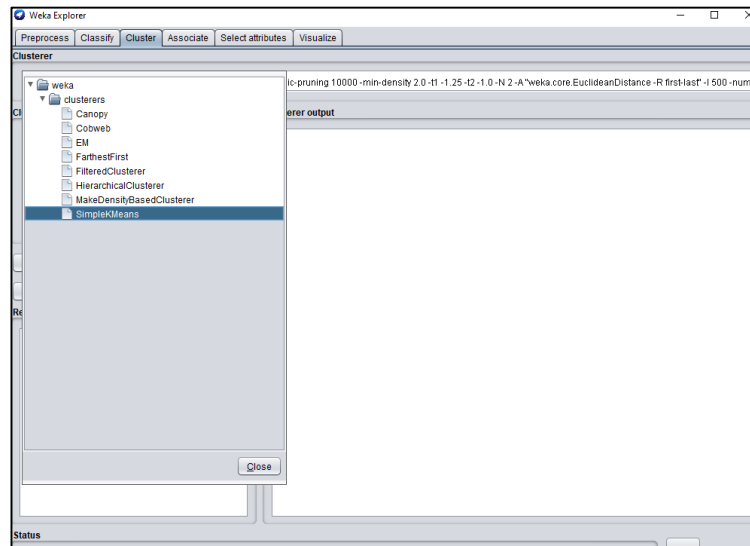


Figura 22. Selección del algoritmo
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- A continuación se configura el algoritmo en cuanto al número de iteraciones, la función de distancia, numero de clusters y el método de inicialización.

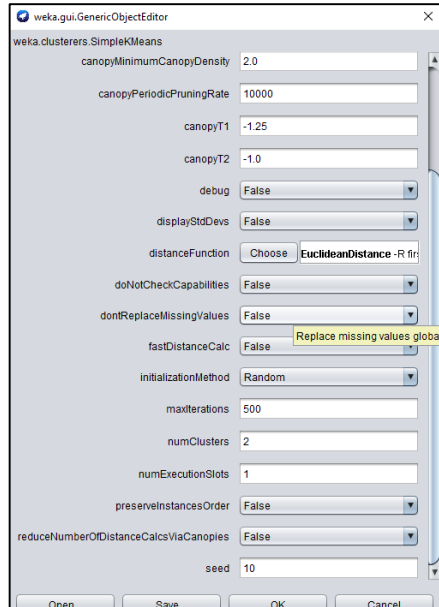


Figura 23. Configuración del algoritmo K-means
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Una vez configurado el algoritmo, se procede a ejecutar el análisis.

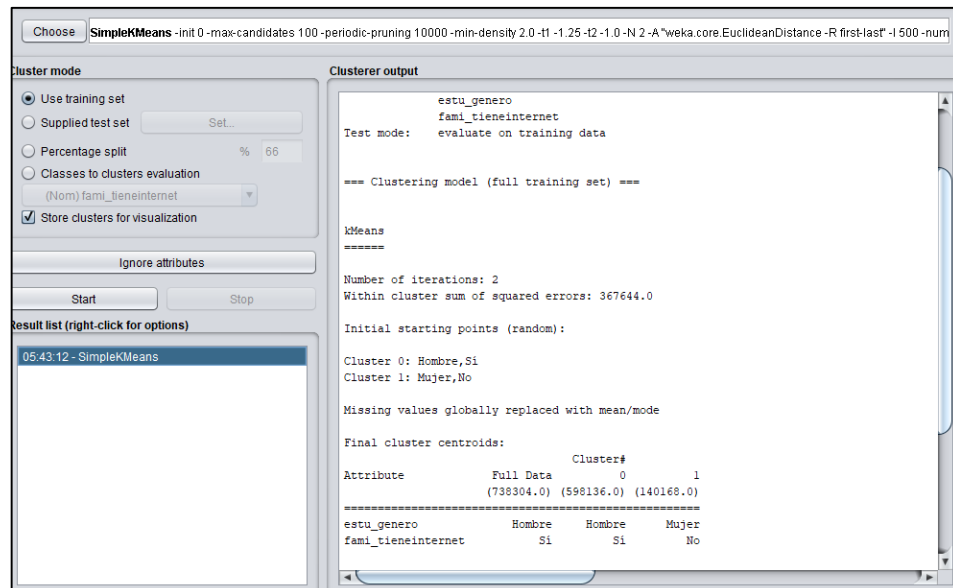


Figura 24. Ejecución del algoritmo K-Means
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Dicho proceso se realizó para cada una de las variables reportadas en el archivo CSV, de esta manera se seleccionaron las que más influencia tuvieron sobre la brecha de género.

3.2. Selección de indicadores

Por ejemplo, en el caso de educación de la madre se pudo evidenciar que para el año 2017 el 74% de las madres de los niños alcanzaron a culminar el bachiller mientras que el 26% de las madres de las niñas termino la primaria.

Attribute	Full Data	Cluster#	
		0	1
	(737044.0)	(542258.0)	(194786.0)
estu_genero	Hombre	Hombre	Mujer
fami_educacionmadre	Bachillerato	Bachillerato	Primaria

Figura 25. Ejemplo de clasificación “Educación madre”

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

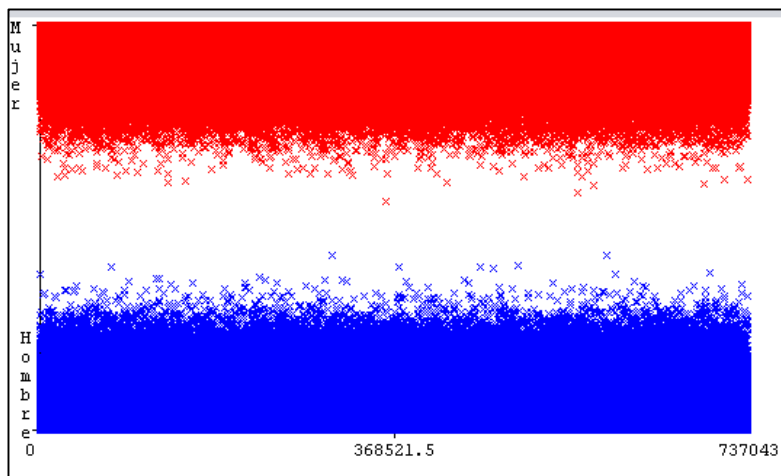


Figura 26. Clusters de “Educación madre”

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

A continuación se muestran las variables analizadas por prueba y los indicadores seleccionados.

Tabla 5. Cantidad de variables analizadas por prueba

Prueba		Saber 3		Saber 5		Saber 9	
Años	Tipo	Variables Analizadas	Variables Seleccionadas	Variables Analizadas	Variables Seleccionadas	Variables Analizadas	Variables Seleccionadas
< 2016	Puntajes	33	3	47	3	47	3
	Socioeconomicos	0		54	5	54	5
2017 >	Puntajes	58	3	58	7	58	7
	Socioeconomicos						

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para la prueba saber tercero se tuvieron en cuenta las siguientes variables:

- Genero del estudiante
- Puntaje en matemáticas
- Puntaje en lenguaje

Para las pruebas saber quinto y noveno se tuvieron en cuenta las siguientes variables:

- Genero del estudiante
- Educación de la madre
- Educación del padre
- Acceso a internet
- Acceso a computador
- Acceso a consola de video juegos
- Puntaje en matemáticas
- Puntaje en lenguaje

4. ALGORITMO DE ANALISIS PREDICTIVO

Para el análisis predictivo se implementó el algoritmo J48 con la ayuda de la herramienta WEKA, inicialmente se analizó el árbol generado y las matrices de

confusión para determinar la validez del modelo, para este proceso se tuvieron en cuenta los siguientes pasos:

- Consultar en la base de datos, información relacionada con el género, indicador y puntaje de algún núcleo temático.
- Seleccionar de la herramienta weka la opción “Classify”, posteriormente seleccionar “trees” y “J48”

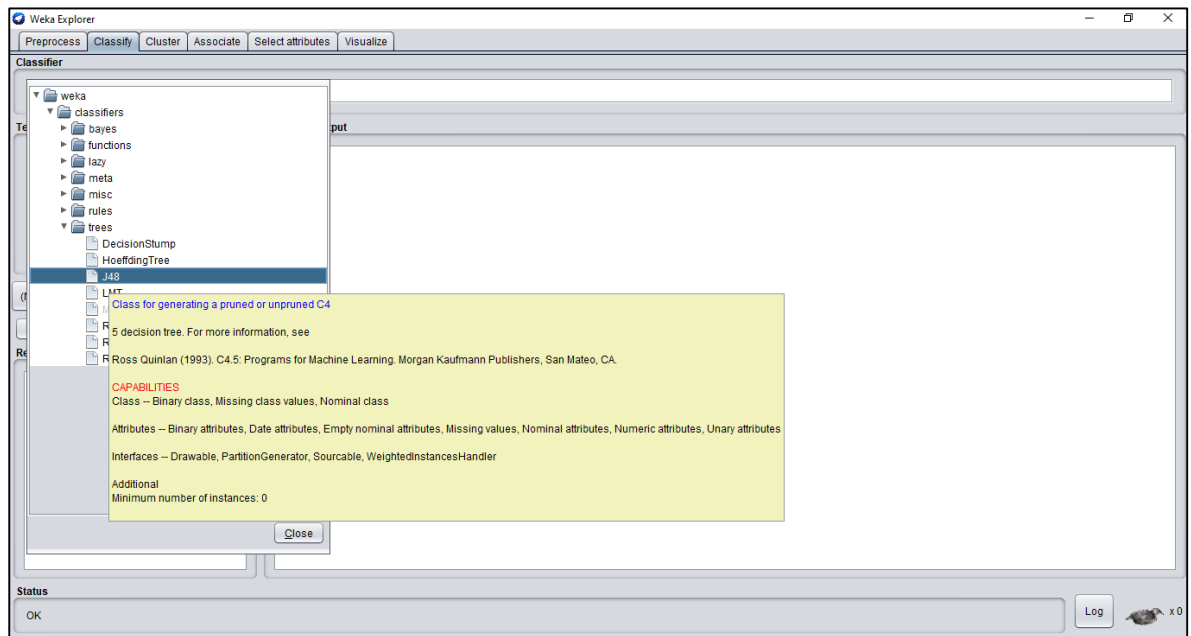


Figura 27. Selección del algoritmo J48
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

- Una vez iniciado el proceso de analisis, la herramienta mostrara el arbol de decision de la siguiente manera:

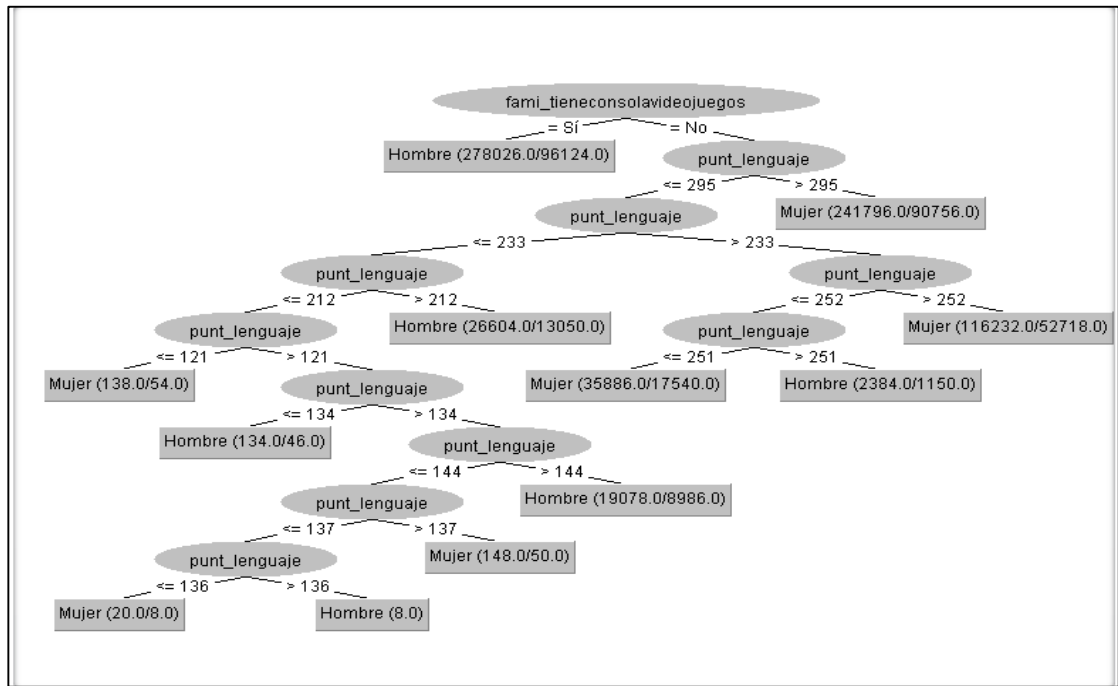


Figura 28. Árbol generado con algoritmo J48
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Una vez generado el árbol el aplicativo procede a extraer las reglas generadas y determinar el porcentaje de clasificación para dicha regla.

4.1. Validación algoritmo predictivo

Para determinar la validez del algoritmo, se tomó una muestra de 50 registros y se realizó la respectiva comparación entre el algoritmo J48 y los algoritmos de predicción RandomForest, RandomTree y REPTree, obteniendo de esta manera los siguientes resultados:

4.1.1. Prueba con indicador de consola de videojuegos

- Porcentaje de clasificación correcta

```
Analysing: Percent_correct
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 09:12

Dataset          (1) trees.J4 | (2) trees (3) trees (4) trees
-----
QueryResult      (100)  90.00 |  79.80 *  80.20 *  89.40
-----
                  (v/ /*) |  (0/0/1)  (0/0/1)  (0/1/0)

Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 2 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 29. Porcentaje de clasificación correcta

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En este resultado se pudo evidenciar que el algoritmo J48 presenta el mayor porcentaje de clasificación correcta, seguido del algoritmo REPTree quien presenta una diferencia de 0.6% frente al primero.

- **Porcentaje de clasificación incorrecta**

```
Analysing: Percent_incorrect
Datasets: 1
Resultsets: 4
Confidence: 1.0 (two tailed)
Sorted by: -
Date: 30/10/20 09:15
```

Dataset	(1) trees.J4	(2) trees	(3) trees	(4) trees
QueryResult	(100) 10.00	20.20 v	19.80 v	10.60 v
	(v/ /*)	(1/0/0)	(1/0/0)	(1/0/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 2 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 30. Porcentaje de clasificación incorrecta
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En el porcentaje de clasificación incorrecta encontramos al algoritmo J48 con el menor porcentaje de error siendo este equivalente al 10%, seguido de REPTree con un 10.6%.

- **Error absoluto medio**

```
Analysing: Mean_absolute_error
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 09:18
```

Dataset	(1) trees.J	(2) tree	(3) tree	(4) tree
QueryResult	(100) 0.18	0.19	0.20	0.19
	(v/ /*)	(0/1/0)	(0/1/0)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 2 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 31. Error absoluto medio.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto al error absoluto medio de los resultados se observa que el algoritmo J48 presenta un 18% de error, siendo este el menor de los cuatro algoritmos analizados.

4.1.2. Prueba con indicador educación de la madre - Porcentaje de clasificación correcta

```
Analysing: Percent_correct
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 09:55
```

Dataset	(1) trees.J48	(2) trees	(3) trees	(4) trees
QueryResult	(100) 90.00	84.60	84.80	89.40
	(v/ /*)	(0/1/0)	(0/1/0)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 32. Porcentaje de clasificación correcta
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para la segunda prueba se puede evidenciar que el algoritmo J48 sigue presentando el mayor porcentaje de clasificación correcta, seguido del algoritmo REPTree quien presenta una diferencia de 0.6% frente al primero, posteriormente RandomTree con un 84.6% de clasificación correcta y por ultimo RandomForest con un 84.8%.

- Porcentaje de clasificación incorrecta

```
Analysing: Percent_incorrect
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 09:57
```

Dataset	(1) trees.J4	(2) trees	(3) trees	(4) trees
QueryResult	(100) 10.00	15.40	15.20	10.60
	(v/ /*)	(0/1/0)	(0/1/0)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 33. Porcentaje de clasificación incorrecta
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para la segunda prueba se puede evidenciar que el algoritmo J48 lidera con un 10% de clasificación incorrecta seguido por REPTree, RandomTree y RandomForest.

- Error absoluto medio

```
Analysing: Mean_absolute_error
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 09:57
```

Dataset	(1) trees.J	(2) tree	(3) tree	(4) tree
QueryResult	(100) 0.18	0.18	0.15	0.18
	(v/ /*)	(0/1/0)	(0/1/0)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 34. Error absoluto medio
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para esta segunda prueba se evidencia que el error es del 18% para los algoritmos: J48, REPTree y RandomForest, mientras que para RandomTree es del 15% siendo este el mejor resultado de esta prueba.

4.1.3. Prueba con indicador de internet - Porcentaje de clasificación correcta

```
Analysing: Percent_correct
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 10:00
```

Dataset	(1) trees.J48	(2) trees	(3) trees	(4) trees
QueryResult	(100) 90.00	79.20 *	78.40 *	90.00
	(v/ /*)	(0/0/1)	(0/0/1)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 35. Porcentaje de clasificación correcta
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para esta última prueba se evidencia que el algoritmo J48 presenta las mejores probabilidades de clasificación correcta, con respecto a las pruebas anteriores se demostró que este algoritmo presenta en promedio un 90% de clasificación correcta, dando así a demostrar su gran eficiencia en el análisis predictivo echo. Asimismo se reconoce al algoritmo REPTree como el segundo algoritmo más efectivo y eficiente durante los procesos predictivos.

- Porcentaje de clasificación incorrecta

```
Analysing: Percent_incorrect
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 10:07
```

Dataset	(1) trees.J4	(2) trees	(3) trees	(4) trees
QueryResult	(100) 10.00	20.80 v	21.60 v	10.00
	(v/ /*)	(1/0/0)	(1/0/0)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 36. Porcentaje de clasificación incorrecta
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para esta última prueba de clasificación incorrecta se evidencia que el algoritmo J48 y REPTree presentan los menores puntajes, dando así a demostrar su eficiencia en el proceso de clasificación de información y generación de reglas predictivas.

- Error absoluto medio

```
Analysing: Mean_absolute_error
Datasets: 1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 30/10/20 10:08
```

Dataset	(1) trees.J	(2) tree	(3) tree	(4) tree
QueryResult	(100) 0.18	0.19	0.22	0.18
	(v/ /*)	(0/1/0)	(0/1/0)	(0/1/0)

```
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428740
(3) trees.RandomTree '-K 0 -M 1.0 -V 0.001 -S 1' -9051119597407395800
(4) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -9216785998198681600
```

Figura 37. Error absoluto medio.
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto al error absoluto medio y en comparación con las pruebas anteriores se evidencia que el algoritmo J48 presenta en promedio presenta un error del 18%, siendo este el menor de todos los resultados.

Analizando todas las pruebas realizadas se puede concluir que el algoritmo J48 presenta mayor precisión sobre los datos y adicionalmente, genera un árbol con resultados eficientes, lo cual corrobora el gran uso que tiene dentro de la minería de datos especialmente en el campo de la educación.

5. DOCUMENTACIÓN DEL SOFTWARE

5.1. Plan del proyecto

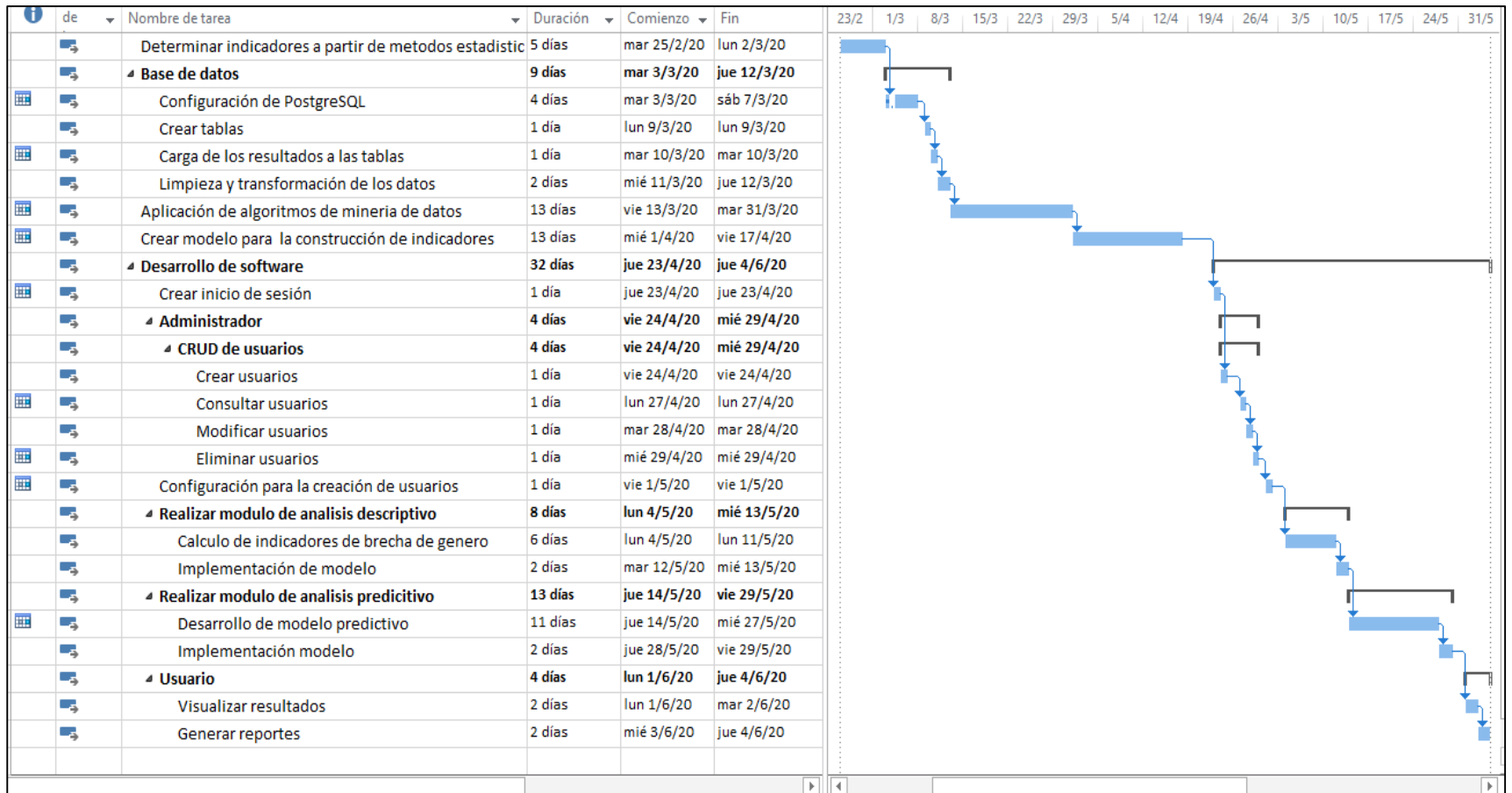


Figura 38. Cronograma de desarrollo de proyecto
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.2. Determinación de requerimientos

Número de requisito	RF 1	
Nombre de requisito	Inicio de sesión	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El usuario podrá acceder a la plataforma a través de un Login, se le pedirá usuario y contraseña para el ingreso.	
Prioridad del requisito	<input type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 2	
Nombre de requisito	Creación usuarios	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	E El administrador podrá diligenciar un formulario con la información pertinente para realizar el registro de nuevos usuarios al aplicativo web. La información se le enviará al correo (usuario y contraseña).	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 3	
Nombre de requisito	Consulta usuarios	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El administrador podrá visualizar el listado de usuarios existentes en el sistema.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 4	
Nombre de requisito	Modificación usuarios	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El administrador podrá actualizar y/o modificar la información de los usuarios, para ello deberá consultar y seleccionar el usuario que desea modificar.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 5	
Nombre de requisito	Eliminación usuarios	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El administrador podrá eliminar los usuarios, consultando previamente y seleccionando el usuario que desea eliminar. Antes de finalizar la operación saldrá una alerta de confirmación.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 6	
Nombre de requisito	Actualización datos de descarga	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El administrador se encargará de actualizar los datos de acceso (URL, usuario y contraseña) del repositorio donde están alojados los datos de las pruebas por parte del ICFES.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 7	
Nombre de requisito	Extracción de los datos	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El aplicativo se encargará de la extracción de los datos disponibles de las pruebas. De forma automática deberá descargar del repositorio del ICFES los archivos necesarios para su posterior análisis, estos vienen en formato “.zip” y al descomprimirlos queda un archivo “.txt” que contiene todos los datos de las pruebas según su nivel.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 8	
Nombre de requisito	Limpieza de los datos	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El aplicativo tendrá la función de limpiar los datos innecesarios (datos nulos y variables que no serán tenidas en cuenta) y asimismo hacer un balanceo de estos para que no se presenten errores durante el cálculo de los indicadores. El aplicativo se encargará de separar la información contenida en el archivo .txt por un carácter especial.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 9	
Nombre de requisito	Carga de los datos	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el aplicativo cargará los datos ya listos para su análisis a la base de datos.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 10	
Nombre de requisito	Cálculo de los indicadores de brecha de género	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El aplicativo se encargará de calcular los indicadores de brecha de género implementando técnicas de minería de datos tales como arboles de decisión y reglas de asociación, con respecto a las variables previamente elegidas para su evaluación.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 11	
Nombre de requisito	Análisis predictivo	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El aplicativo se encargará de calcular y determinar si existe relación entre los núcleos temáticos en común, desde las pruebas de la educación básica hasta las del nivel superior, para así poder predecir su comportamiento. Esto se realizará por medio de algoritmos de árboles de decisión, los cuales nos permitirán conocer la conexión, si es que existe, entre pruebas de cada una de las áreas evaluadas con respecto a la brecha de género.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 12	
Nombre de requisito	Visualización de los resultados	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El usuario podrá visualizar los resultados de forma gráfica y estadística filtrados por el año que desee obtener información y/o por los indicadores hallados durante la investigación y análisis.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 13	
Nombre de requisito	Reportes	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	El aplicativo le permitirá al usuario visualizar y descargar los reportes de resultados de los análisis previamente realizados, incluyendo gráficas y estadísticas. El formato de descarga será PDF.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 14	
Nombre de requisito	Cargar pruebas	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el usuario subirá las pruebas comprimidas en un archivo “.ZIP” al aplicativo, una vez subido se guardara el archivo y se realizara el registro de carga.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 15	
Nombre de requisito	Consultar resultados de los indicadores	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad se le mostrara al usuario los resultados hallados con el algoritmo de minería de datos. Para ello el usuario seleccionara el tipo de prueba y el indicador a analizar.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 16	
Nombre de requisito	Consultar resultados del análisis predictivo	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el aplicativo mostrara al usuario los resultados hallados durante el análisis predictivo.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 17	
Nombre de requisito	Consultar listado de pruebas	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el aplicativo mostrara al usuario el listado de pruebas cargadas a la base de datos.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 18	
Nombre de requisito	Consultar registro de mujeres por prueba	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el aplicativo mostrara al usuario la cantidad de mujeres registradas por prueba.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 19	
Nombre de requisito	Consultar cantidad de registros por prueba	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el aplicativo mostrara al administrador la cantidad de registros existentes por prueba.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 20	
Nombre de requisito	Cerrar Sesión	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el usuario podrá terminar su sesión dentro del aplicativo	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

5.2.1. Modificación de requerimientos

Número de requisito	RF 7	
Nombre de requisito	Extracción de los datos	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el sistema se encargara de descomprimir los archivos .zip y de guardar las pruebas en una carpeta específica.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 9	
Nombre de requisito	Carga de los datos	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad el gestor de base de datos se encargara de cargar las pruebas descomprimidas a la base de datos. En caso de que existan inconvenientes con el archivo el aplicativo se encargara de notificar al administrador sobre el error	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

Número de requisito	RF 8	
Nombre de requisito	Limpieza de los datos	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	En esta actividad la base de datos tendrá la función de eliminar las columnas innecesarias y los datos nulos existentes en las pruebas.	
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/Opcional	

5.3. Especificación del diseño

5.3.1. Modelo Entidad Relación

5.3.1.1. Modelo entidad relación de la base de datos ingappe

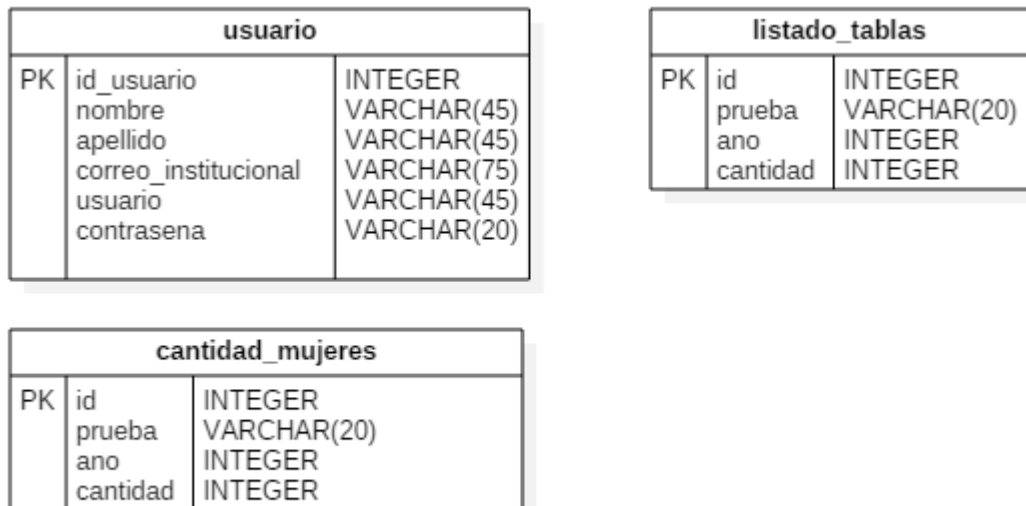


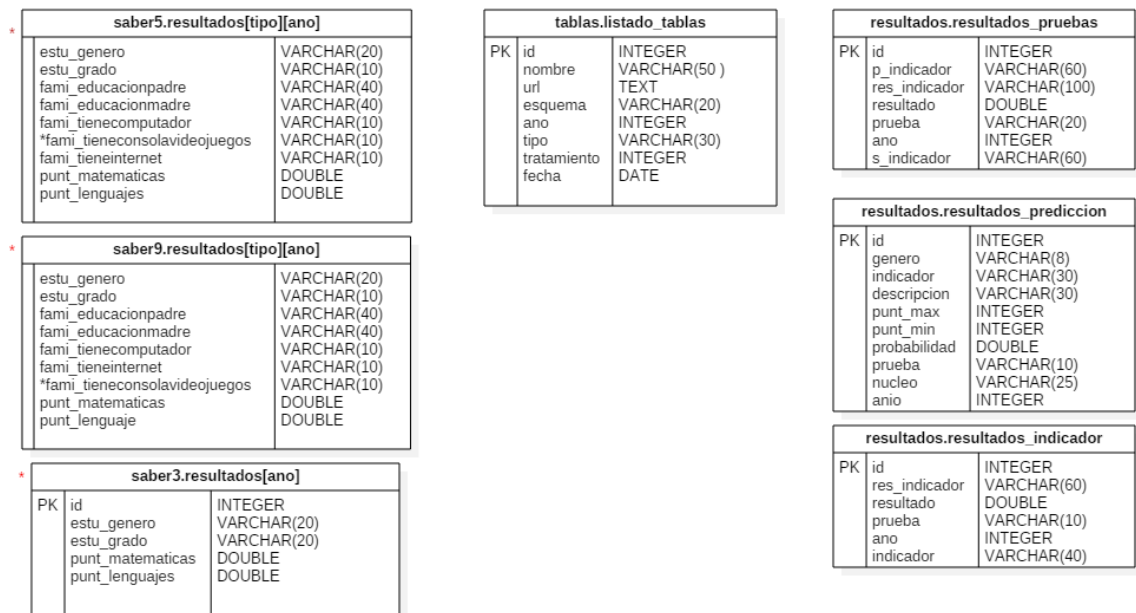
Figura 39. Modelo Entidad Relación, Base de datos Ingappe
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 6. Descripción del Modelo entidad relación de la base de datos ingagpe

Tabla	Descripción
usuario	La tabla usuario almacena la información básica de las personas interesadas en acceder al aplicativo.
cantidad_mujeres	La tabla cantidad_mujeres almacena la cantidad de mujeres registradas por prueba y año respectivamente.
listado_tablas	La tabla listado_tablas almacena la cantidad de registros de las pruebas recolectadas por el aplicativo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.1.2. Modelo entidad relación de la base de datos resultados_basico



Observaciones:

- Las tablas de los esquemas saber 3, 5 y 9 se crearan de acuerdo a los años y tipo de resultados (economicos o puntajes) cargados en el aplicativo
- La variable fami_tieneconsolavideojuegos será cargada de acuerdo al año, ya que existen pruebas que no contienen información al respecto

Figura 40. Modelo Entidad Relación Base de datos resultados básico
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 7. Descripción del modelo entidad relación de la base de datos resultados_basico

Tabla	Descripción
saber3.resultados[ano]	Las tablas saber3.resultados son aquellas donde se almacenan datos relacionados con los resultados de las pruebas saber 3, estas tablas se crearan una vez se carguen pruebas al aplicativo. La creación dependerá del año que se cargue.
saber5.resultados[tipo][ano]	Las tablas saber5.resultados son aquellas donde se almacenan datos relacionados con los resultados de las pruebas saber 5, estas tablas se crearan una vez se carguen pruebas al aplicativo. La creación dependerá del año y el tipo de datos (puntajes o socioeconómicos) que se carguen.
saber9.resultados[tipo][ano]	Las tablas saber9.resultados son aquella donde se almacenan datos relacionados con los resultados de las pruebas saber 9, estas tablas se crearan una vez se carguen pruebas al aplicativo. La creación dependerá del año y el tipo de datos (puntajes o socioeconómicos) que se carguen.
tablas.listado_tabla	La tabla listado_tabla almacena la información relacionada con las pruebas recolectadas por el aplicativo
resultados.resultados_prueba	La tabla almacena los resultados hallados en el aplicativo de acuerdo al tipo prueba
resultados.resultados_indicador	La tabla almacena los resultados analizados en el aplicativo por cada indicador.
resultados.resultados_prediccion	La tabla almacena los resultados hallados en el análisis predictivo

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

5.3.2. Roles

Los roles son el papel que cumple cada persona a partir de sus funciones sobre el aplicativo, para así llevar a cabo un orden y ciertas responsabilidades.

5.3.2.1. Administrador

Este usuario se encarga de crear, consultar, modificar y eliminar usuarios del aplicativo, adicionalmente tiene la responsabilidad de cargar las pruebas al aplicativo y es el único usuario que puede visualizar el listado de pruebas disponibles.

5.3.2.2. Usuario

Este rol se encarga de consultar los resultados hallados por cada prueba, indicador y núcleo temático.

5.3.2.3. Aplicativo

Este rol se encarga de implementar el modelo de minería de datos y de generar resultados y reportes (gráficos y estadísticas) para los usuarios, adicionalmente se encarga de guardar las pruebas en el sistema.

5.3.2.4. Base de datos

Este rol se encarga de realizar el proceso de carga, transformación y limpieza de los datos por cada prueba cargada.

5.3.2.5. Sistema

Este rol se encarga de descomprimir los archivos .zip y guardar los documentos de texto en una carpeta

5.3.3. Diagramas de casos de uso

5.3.3.1. Casos de uso: usuario

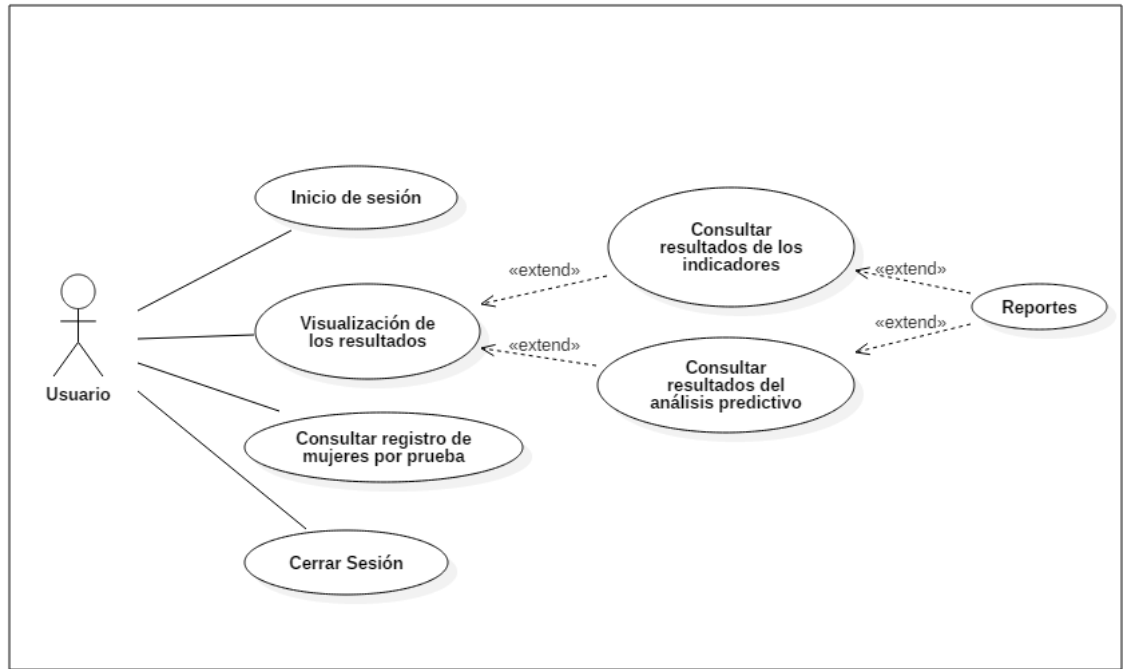


Figura 41. Casos de uso: Usuario

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 8. Descripción de caso de uso inicio de sesión

CASO DE USO	
Nombre	Inicio de sesión
Actores	Usuario
Función	Acceder al módulo de resultados
Descripción	El aplicativo validara el acceso al módulo de resultados, con el usuario y la contraseña dadas por el administrador.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 9. Descripción de caso de uso visualización de los resultados

CASO DE USO	
Nombre	Visualización de los resultados
Actores	Usuario
Función	Visualizar los resultados hallados durante los análisis hechos.
Descripción	Obtener y visualizar los resultados hallados durante el análisis descriptivo y predictivo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 10. Descripción de caso de uso consultar resultados de los indicadores

CASO DE USO	
Nombre	Consultar resultados de los indicadores
Actores	Usuario
Función	Visualizar los resultados hallados durante el análisis descriptivo
Descripción	Obtener y visualizar los resultados hallados durante el análisis descriptivo, de acuerdo al indicador, año y tipo de prueba.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 11. Descripción de caso de uso consultar resultados del análisis predictivo

CASO DE USO	
Nombre	Consultar resultados del análisis predictivo
Actores	Usuario
Función	Visualizar los resultados hallados durante el análisis predictivo
Descripción	Obtener y visualizar los resultados hallados durante el análisis predictivo según el tipo de prueba y el núcleo temático.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 12. Descripción de caso de uso reportes

CASO DE USO	
Nombre	Reportes
Actores	Usuario
Función	Obtener reportes por cada resultado encontrado.
Descripción	El usuario podrá visualizar y descargar reportes de los resultados obtenidos, allí obtendrá información más detallada y graficas sobre los análisis hechos.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 13. Descripción de caso de uso consultar registro de mujeres por prueba

CASO DE USO	
Nombre	Consultar registro de mujeres por prueba
Actores	Usuario
Función	Consultar cantidad de mujeres registradas por prueba.
Descripción	El usuario visualizara gráficamente y estadísticamente la cantidad de mujeres inscritas en cada una de las pruebas reportadas en el aplicativo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 14. Descripción de caso de uso cerrar sesión

CASO DE USO	
Nombre	Cerrar Sesión
Actores	Usuario
Función	Cerrar módulo de los resultados
Descripción	El usuario podrá cerrar sesión del módulo de análisis descriptivo y predictivo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.3.2. Casos de uso: administrador

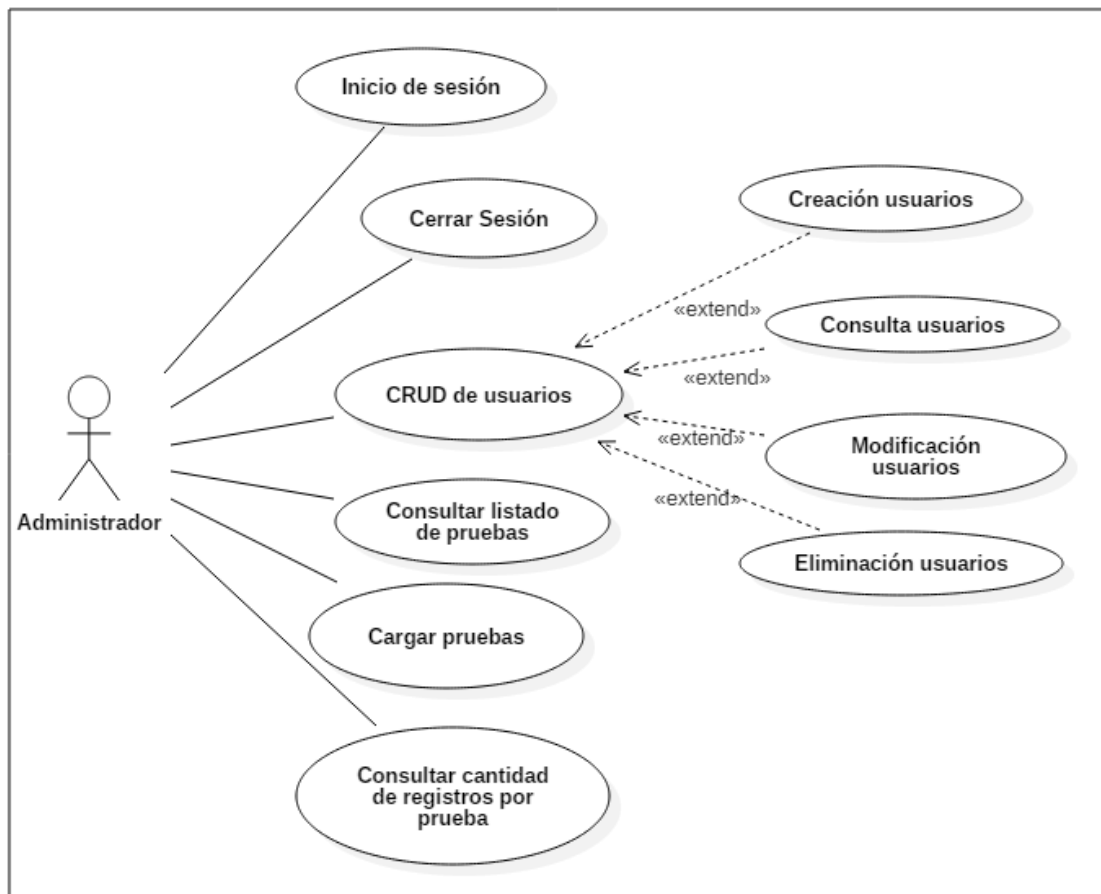


Figura 42. Casos de uso: Administrador
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 15. Descripción de caso de uso inicio de sesión

CASO DE USO	
Nombre	Inicio de sesión
Actores	Administrador
Función	Acceder al módulo administrativo del aplicativo
Descripción	El aplicativo validara el acceso al módulo administrativo, con el usuario y la contraseña.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 16. Descripción de caso de uso cerrar sesión

CASO DE USO	
Nombre	Cerrar Sesión
Actores	Administrador
Función	Cerrar módulo administrativo del aplicativo
Descripción	Se le permite al administrador cerrar sesión del módulo administrativo del aplicativo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 17. Descripción de caso de uso CRUD de usuarios

CASO DE USO	
Nombre	CRUD de usuarios
Actores	Administrador
Función	Crear, consultar, modificar y eliminar usuarios del aplicativo
Descripción	Se le permite al administrador crear, consultar, modificar y eliminar usuarios o datos del mismo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 18. Descripción de caso de uso creación usuarios

CASO DE USO	
Nombre	Creación usuarios
Actores	Administrador
Función	Crear usuarios
Descripción	El administrador podrá crear usuarios una vez le sea notificado por correo la solicitud de registro.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 19. Descripción de caso de uso consulta usuarios

CASO DE USO	
Nombre	Consulta usuarios
Actores	Administrador
Función	Obtener un listado de usuarios.
Descripción	El administrador podrá visualizar un listado de todos los usuarios registrados en el aplicativo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 20. Descripción de caso de uso modificación usuarios

CASO DE USO	
Nombre	Modificación usuarios
Actores	Administrador
Función	Modificar datos del usuario.
Descripción	El administrador podrá modificar o actualizar datos de los usuarios registrados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 21. Descripción de caso de uso eliminación usuarios

CASO DE USO	
Nombre	Eliminación usuarios
Actores	Administrador
Función	Eliminar los datos registrados del usuario
Descripción	El administrador podrá eliminar toda la información relacionada con el usuario.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 22. Descripción de caso de uso cargar pruebas

CASO DE USO	
Nombre	Cargar pruebas
Actores	Administrador
Función	Cargar pruebas al aplicativo
Descripción	El administrador se encargara de realizar la carga de pruebas al aplicativo, esto dependerá de los años y las pruebas que no han sido cargadas. Para la carga se debe subir un archivo “.zip”, una vez subido el aplicativo se encargara de validar y guardar el archivo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 23. Descripción de caso de uso consultar listado de pruebas

CASO DE USO	
Nombre	Consultar listado de pruebas
Actores	Administrador
Función	Consultar listado de pruebas disponibles
Descripción	El administrador visualizara un listado con todas las pruebas cargadas exitosamente a la base de datos.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 24. Descripción de caso de uso consultar cantidad de registros por prueba

CASO DE USO	
Nombre	Consultar cantidad de registros por prueba
Actores	Administrador
Función	Consultar la cantidad de registros por prueba
Descripción	El administrador visualizara un listado con la cantidad de estudiantes registrados por prueba, allí podrá realizar filtros por prueba o año. Asimismo, se le mostraran graficas de los datos consultados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.3.3. Casos de uso: Base de datos

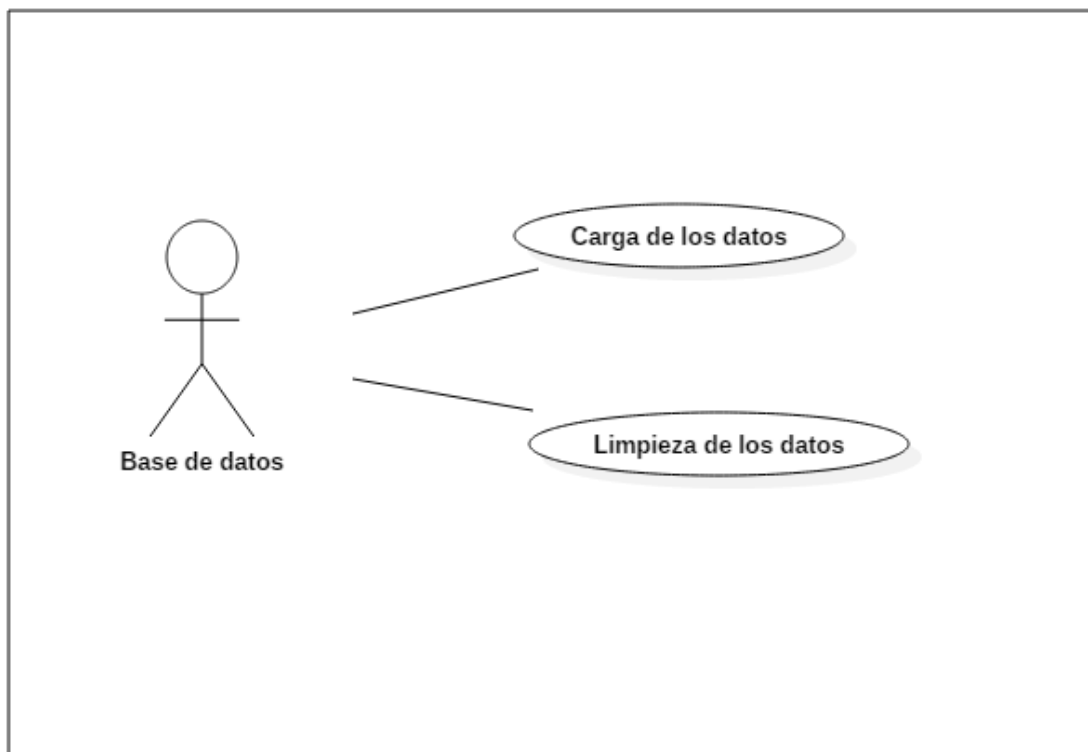


Figura 43. Casos de uso: base de datos

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 25. Descripción de caso de uso carga de los datos

CASO DE USO	
Nombre	Carga de los datos
Actores	Base de datos
Función	Cargar los archivos “csv” o “txt” a la base de datos
Descripción	La base de datos por medio de tareas programadas se encargara de cargar la información de las pruebas a la tabla indicada.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 26. Descripción de caso de uso limpieza de los datos

CASO DE USO	
Nombre	Limpieza de los datos
Actores	Base de datos
Función	Realizar limpieza a las tablas creadas
Descripción	Una vez cargada la información de los archivos a la tabla, la base de datos se encargara de eliminar las columnas innecesarias y los datos nulos.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.3.4. Casos de uso: aplicativo

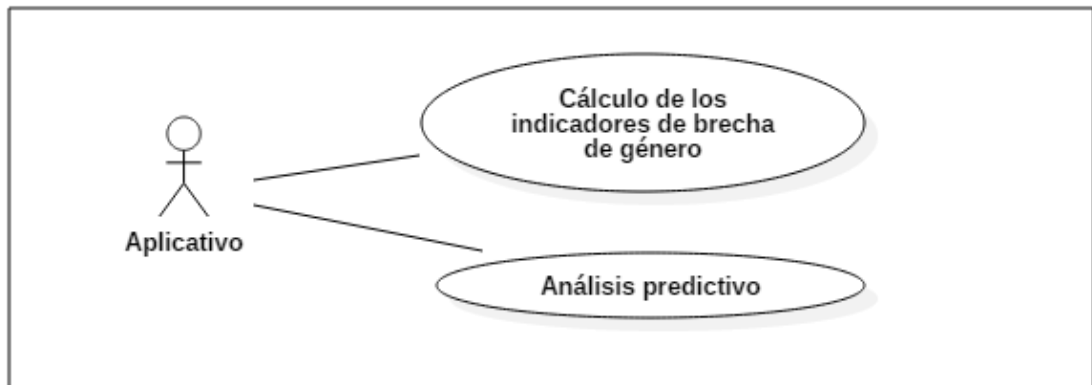


Figura 44. Casos de uso: Aplicativo

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 27. Descripción de caso de uso cálculo de los indicadores de brecha de género

CASO DE USO	
Nombre	Calculo de los indicadores de brecha de genero
Actores	Aplicativo
Función	Realizar el proceso de cálculo de brecha de género por indicador
Descripción	El aplicativo se encargara de hallar resultados descriptivos por medio del algoritmo K-means de la librería de WEKA, allí se obtendrá porcentualmente la influencia del indicador sobre el género

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 28. Descripción de caso de uso análisis predictivo

CASO DE USO	
Nombre	Análisis predictivo
Actores	Aplicativo
Función	Realizar el proceso de análisis predictivo
Descripción	El aplicativo se encargara de hallar resultados predictivos en cuanto a la influencia de los indicadores sobre los núcleos temáticos. Para la implementación de la parte predictiva el aplicativo hace uso del algoritmo J48 de la librería WEKA.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.3.5. Casos de uso: sistema

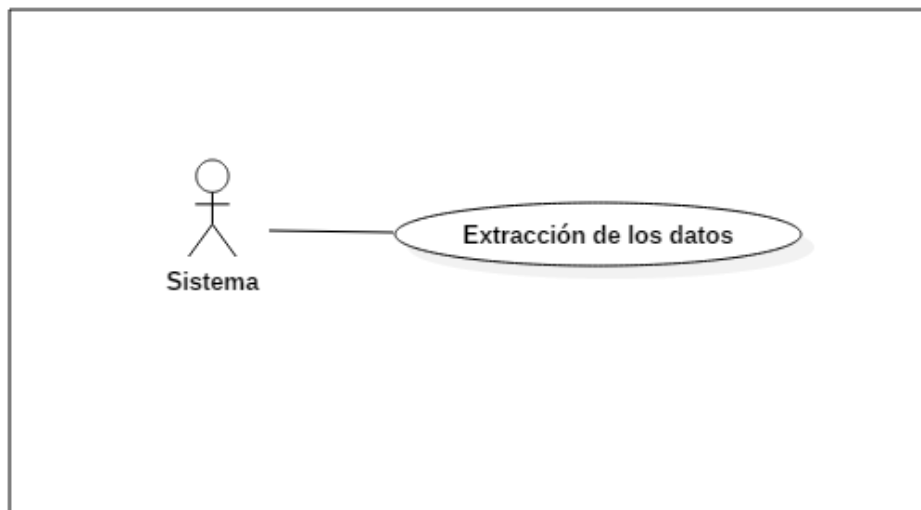


Figura 45. Casos de uso: Sistema

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 29. Descripción de caso de uso extracción de los datos

CASO DE USO	
Nombre	Extracción de los datos
Actores	Sistema
Función	Extraer los documentos de texto del archivo “.zip”
Descripción	Una vez el aplicativo cargue y guarde las pruebas en el servidor, el sistema se encargara por medio de tareas programadas de extraer los documentos del archivo “.zip”.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4. Diagramas de secuencia

El diagrama de secuencia es usado para identificar el comportamiento de cada uno de los roles y la interacción de los mismos frente a cada acción, estos se deben elaborar por cada caso de uso.

5.3.4.1. Diagrama de secuencia inicio de sesión

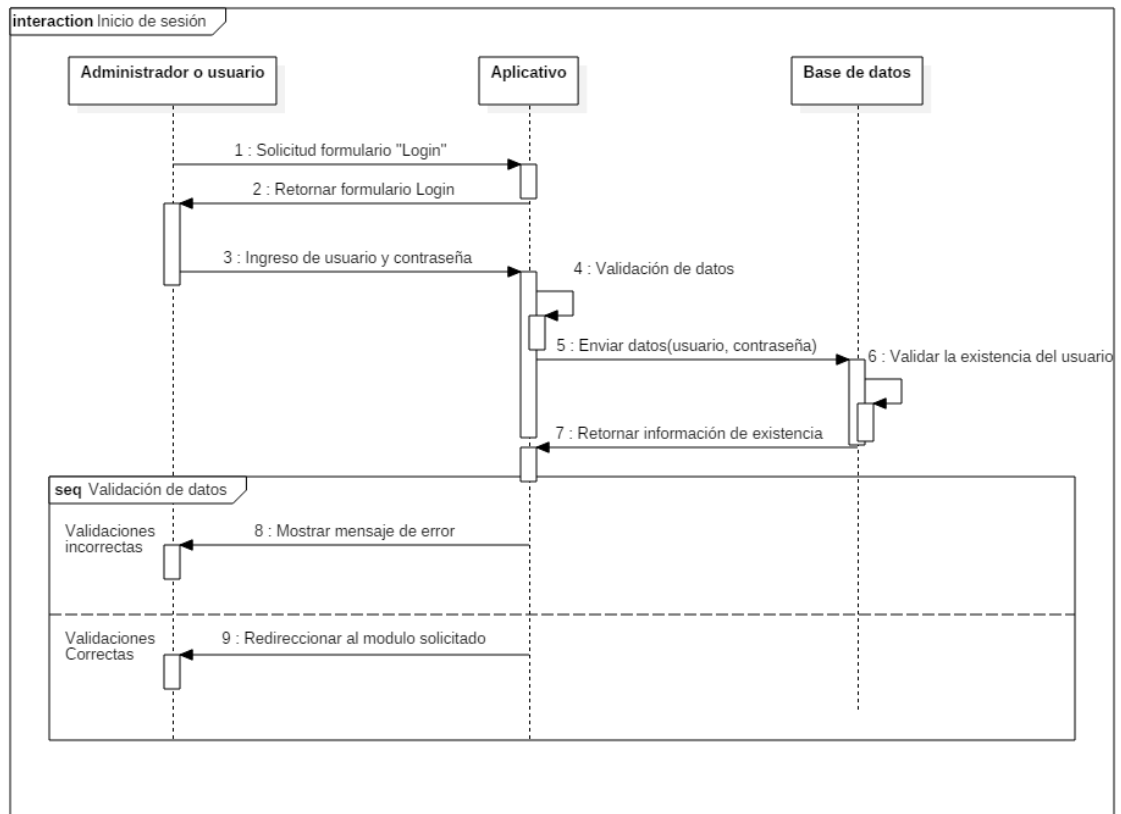


Figura 46. Diagrama de secuencia: inicio de sesión
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 30. Descripción diagrama de secuencia: inicio de sesión

Objeto	Descripción
Administrador o Usuario	El administrador o el usuario tiene una secuencia de actividades, la cual inicia con la solicitud del formulario de "Inicio de Sesión", allí deberá ingresar los datos de usuario y contraseña, una vez se valide la información; si es correcta podrá ingresar al módulo solicitado de lo contrario se le mostrara un mensaje de error.
Aplicativo	El aplicativo se encargara inicialmente de mostrar el formulario solicitado, posteriormente realizara la validación de los datos ingresados al aplicativo y por ultimo redireccionará al módulo solicitado o retornara un mensaje de error
Base de datos	La base de datos se encargara de verificar si el usuario y la contraseña ingresada coinciden con algún registro de la base de datos.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.2. Diagrama de secuencia cerrar sesión

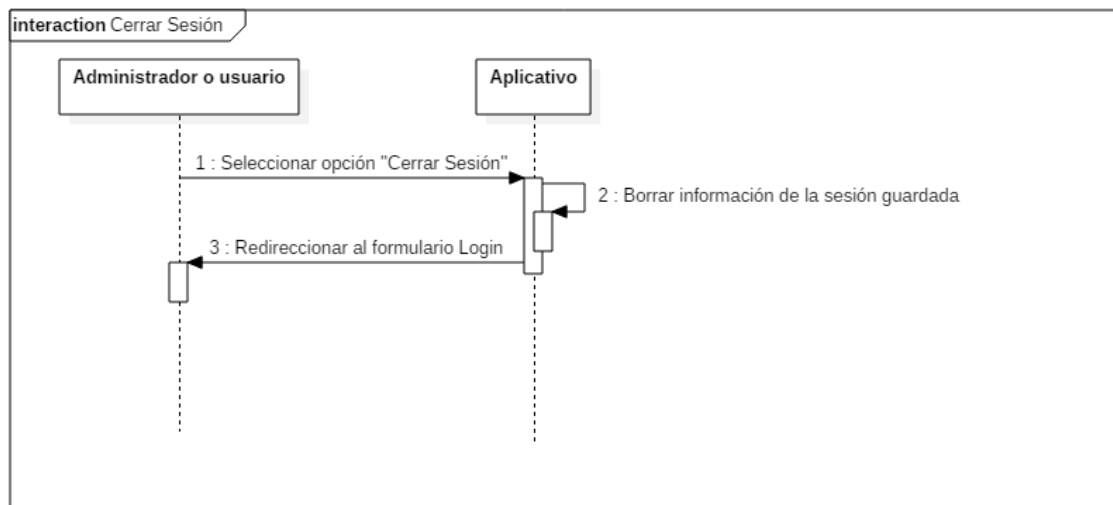


Figura 47. Diagrama de secuencia: cerrar Sesión
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 31. Descripción diagrama de secuencia: cerrar sesión

Objeto	Descripción
Administrador o usuario	El administrador o el usuario tiene una secuencia de actividades, la cual cuenta con una sola acción que es la de "Cerrar sesión"
Aplicativo	El aplicativo se encargara de cerrar la sesión del usuario y borrar la sesión registrada dentro del navegador.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

5.3.4.3. Diagrama de secuencia creación usuarios

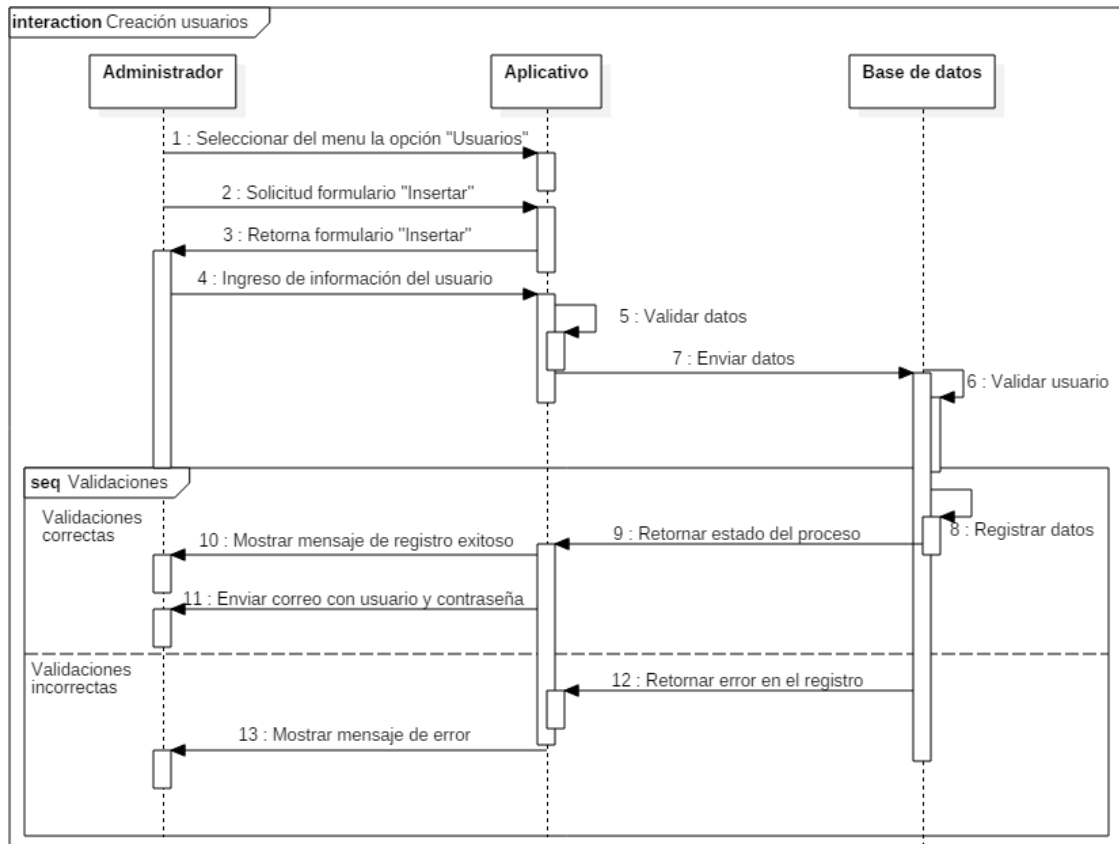


Figura 48. Diagrama de secuencia: creación usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 32. Descripción diagrama de secuencia: creación usuarios

Objeto	Descripción
Administrador	El administrador tiene una secuencia de actividades, la cual inicia con la selección de “Usuarios” del menú principal, posteriormente seleccionara la opción “Insertar”. En el formulario deberá ingresar los datos del usuario a registrar, una vez se valide la información se le mostrara un mensaje indicándole si el proceso fue exitoso o presento inconvenientes.
Aplicativo	El aplicativo se encargara inicialmente de mostrar el formulario solicitado, posteriormente realizara la validación de los datos ingresados al aplicativo y por último, retornara un mensaje indicando si el proceso fue exitoso o no. Adicionalmente retornara un correo a la persona registrada indicando el usuario y la contraseña asignada.
Base de datos	La base de datos se encargara de validar que el usuario no exista y de insertar a la tabla de usuarios la información ingresada de la persona.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.4. Diagrama de secuencia consulta usuarios

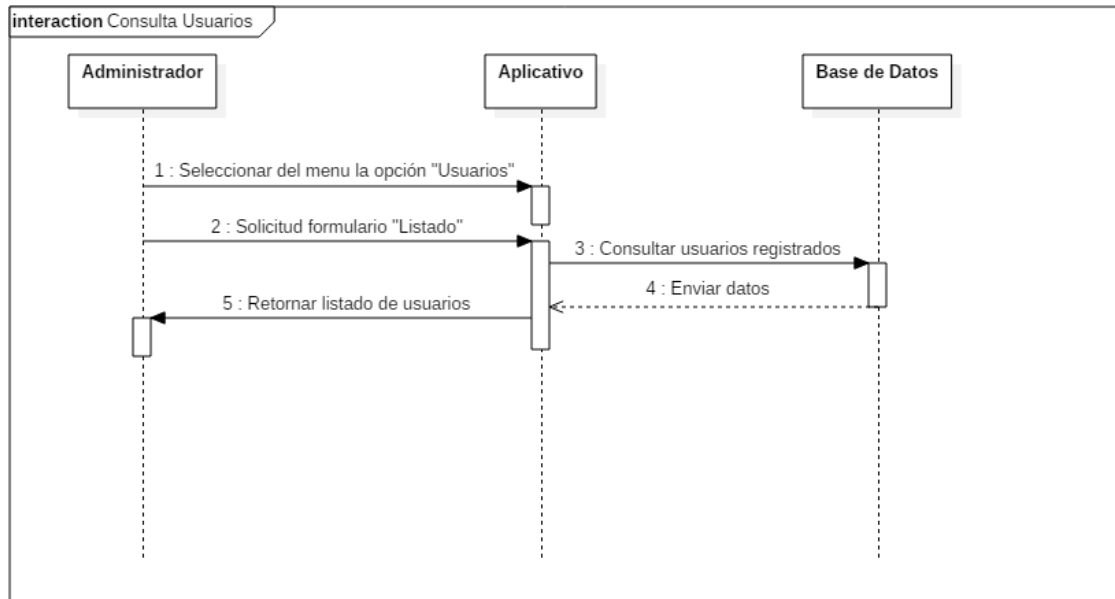


Figura 49. Diagrama de secuencia: consulta usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 33. Descripción diagrama de secuencia: consulta usuarios

Objeto	Descripción
Administrador	El administrador tiene una secuencia de actividades, la cual inicia con la selección de “Usuarios” del menú principal, posteriormente seleccionara la opción “Listado”, después de ello el aplicativo le mostrara un listado con todos los usuarios registrados.
Aplicativo	El aplicativo se encargara inicialmente de generar un formulario con el listado de los usuarios registrados en base de datos.
Base de datos	La base de datos se encargara de retornar la información de los usuarios registrados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.5. Diagrama de secuencia modificación usuarios

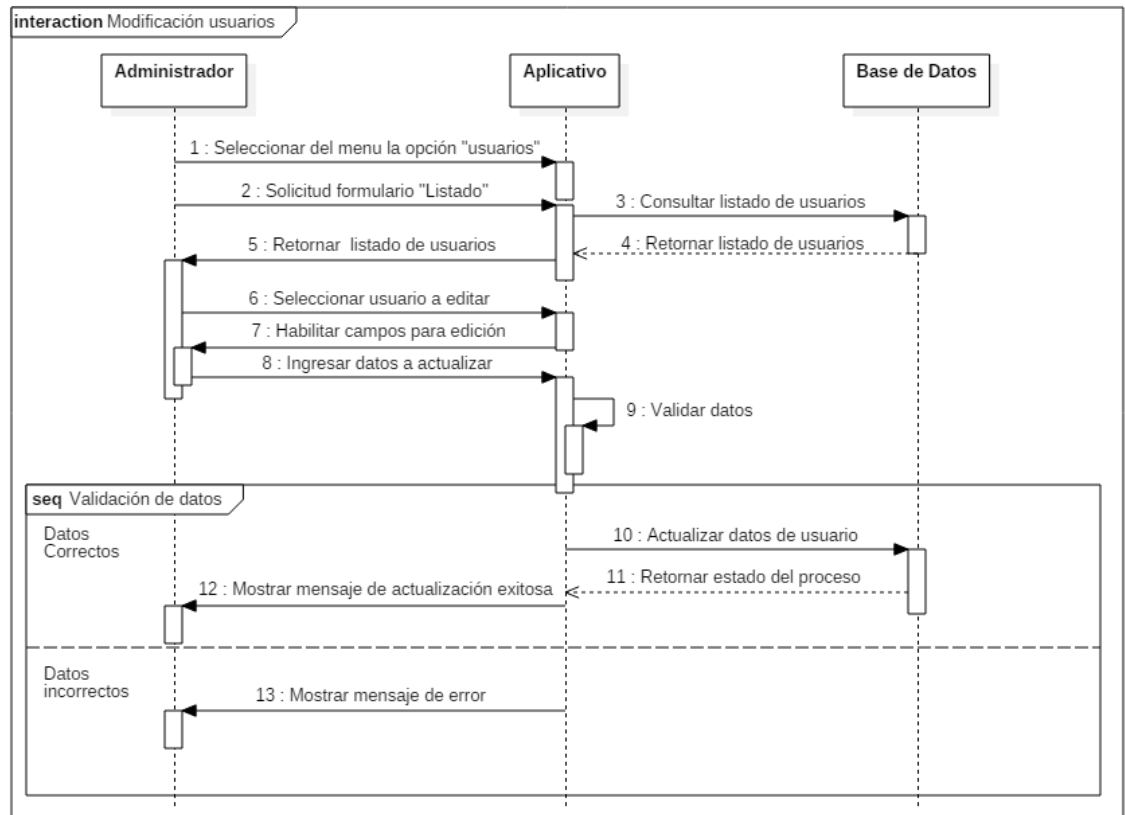


Figura 50. Diagrama de secuencia: modificación usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 34. Descripción diagrama de secuencia: modificación usuarios

Objeto	Descripción
Administrador	El administrador tiene una secuencia de actividades, la cual inicia con la selección de “Usuarios” del menú principal, posteriormente seleccionara la opción “Listado”, a partir del listado de usuarios, seleccionara el usuario a editar e ingresara los datos a modificar, una vez el aplicativo valide la información se le enviara un mensaje indicando el estado del proceso solicitado.
Aplicativo	El aplicativo se encargara inicialmente de generar un formulario con el listado de los usuarios registrados en base de datos, posteriormente se encargara de habilitar los campos para la edición y validar los datos ingresados, una vez realizado dicho proceso se encargara de enviar un mensaje indicando si el proceso fue exitoso.
Base de datos	La base de datos se encargara de consultar los usuarios registrados y de editar los datos del usuario seleccionado.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.6. Diagrama de secuencia eliminación usuarios

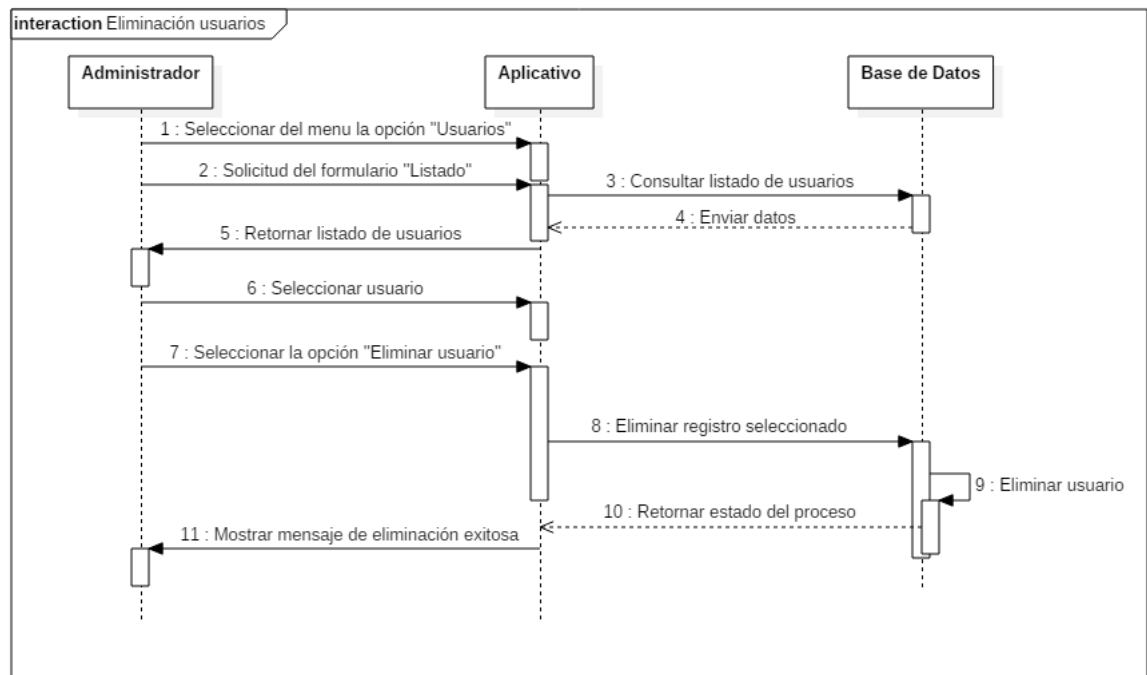


Figura 51. Diagrama de secuencia: eliminación usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 35. Descripción diagrama de secuencia: eliminación usuarios

Objeto	Descripción
Administrador	El administrador tiene una secuencia de actividades, la cual inicia con la selección de "Usuarios" del menú principal posteriormente seleccionara la opción "Listado". A partir del listado de usuarios seleccionara el usuario a eliminar.
Aplicativo	El aplicativo se encargara inicialmente de generar un formulario con el listado de los usuarios registrados en base de datos, posteriormente enviara un mensaje indicando la eliminación del usuario.
Base de datos	La base de datos se encargara de eliminar la información del usuario.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.7. Diagrama de secuencia cargar pruebas

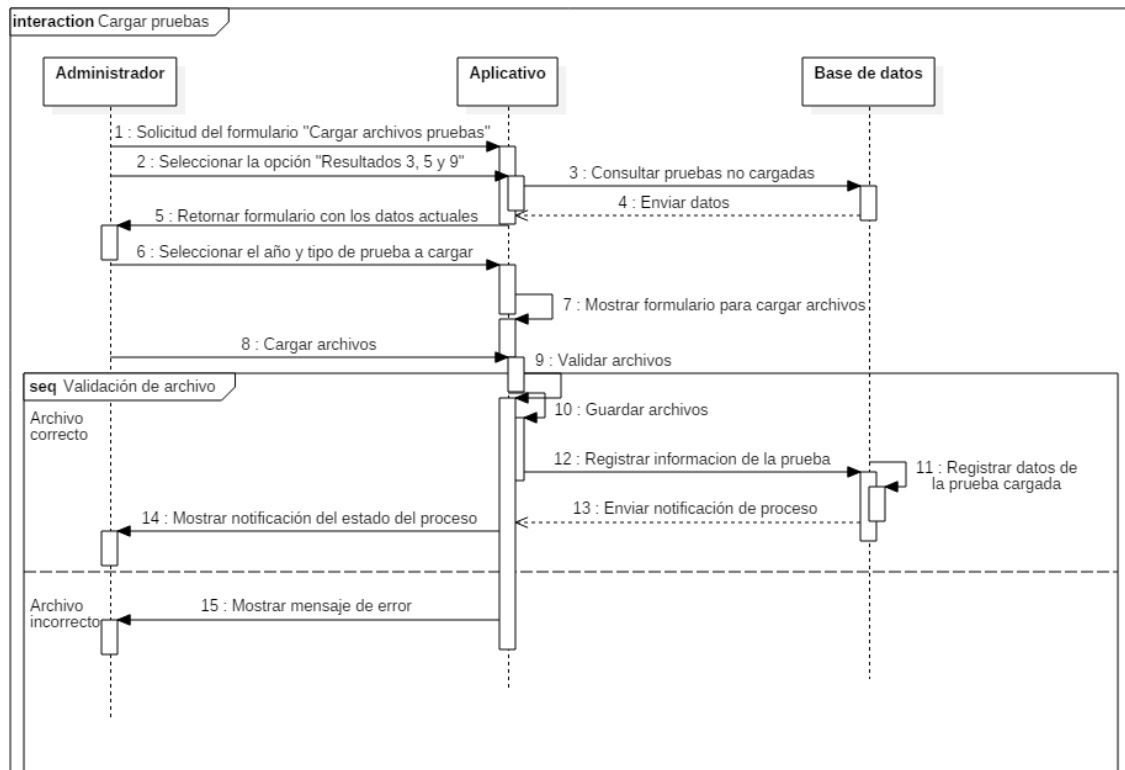


Figura 52. Diagrama de secuencia: cargar pruebas
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 36. Descripción diagrama de secuencia: cargar pruebas

Objeto	Descripción
Administrador	El administrador tiene una secuencia de actividades, la cual inicia con la solicitud del formulario "Cargar archivos pruebas" allí seleccionara "Resultados 3, 5 y 9" posteriormente, deberá seleccionar el año y el tipo de prueba a cargar; una vez seleccionado realizara la carga del archivo o los archivos.
Aplicativo	El aplicativo se encargara inicialmente de generar un formulario con la información de las pruebas que aún no han sido cargadas, posteriormente se encargara de validar el archivo o los archivos cargados por el usuario, una vez validados guardara los archivos en el servidor y por ultimo generara un mensaje indicando el estado del proceso
Base de datos	La base de datos se encargara de registrar los datos de la prueba cargada.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.8. Diagrama de secuencia extracción de los datos

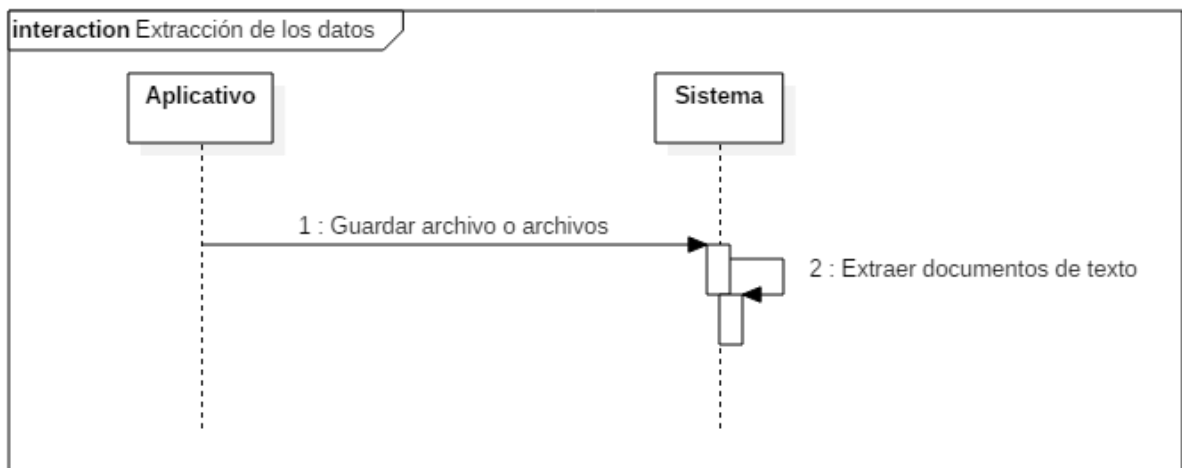


Figura 53. Diagrama de secuencia: extracción de los datos

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 37. Descripción diagrama de secuencia: extracción de los datos

Objeto	Descripción
Aplicativo	El aplicativo se encargara de guardar los archivos “.zip” que fueron cargados previamente por el usuario.
Sistema	El sistema se encargara de descomprimir los archivos y verificar que correspondan al tipo de formato solicitado (txt o csv).

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.9. Diagrama de secuencia carga de los datos

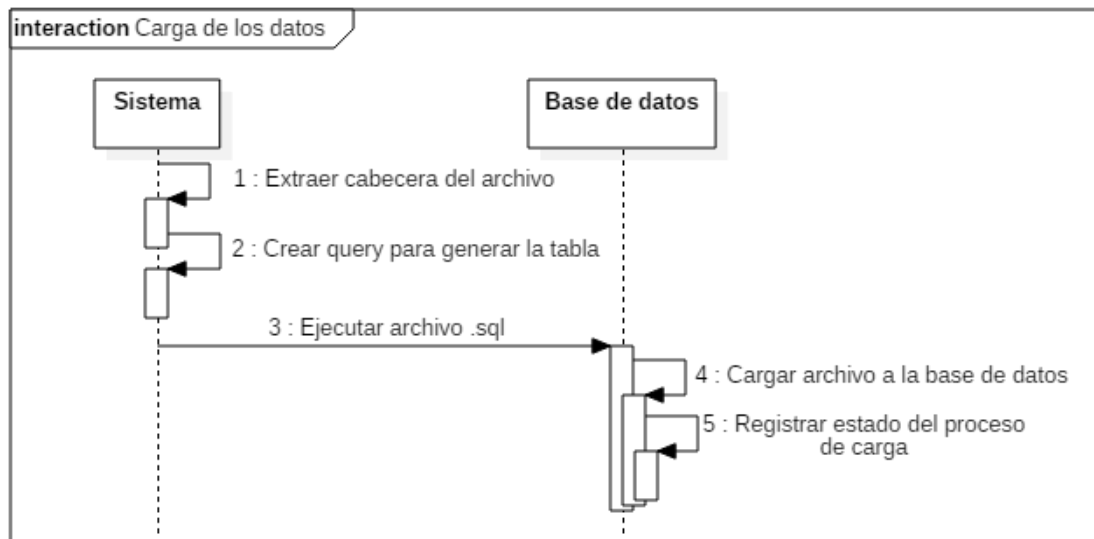


Figura 54. Diagrama de secuencia: carga de los datos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 38. Descripción diagrama de secuencia: carga de los datos

Objeto	Descripción
Sistema	El sistema a partir del archivo cargado, se encargara de extraer la cabecera para definir las columnas de la tabla, adicionalmente se encargara de crear un archivo “.sql” el cual contendrá el query para la creación de la tabla.
Base de datos	La base de datos al ejecutar el archivo “.sql”, se encargara de crear la tabla y posteriormente de copiar toda la información del archivo a la tabla, asimismo, registrara el estado del proceso de carga; esto con el fin de llevar control de los archivos cargados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.10. Diagrama de secuencia limpieza de los datos

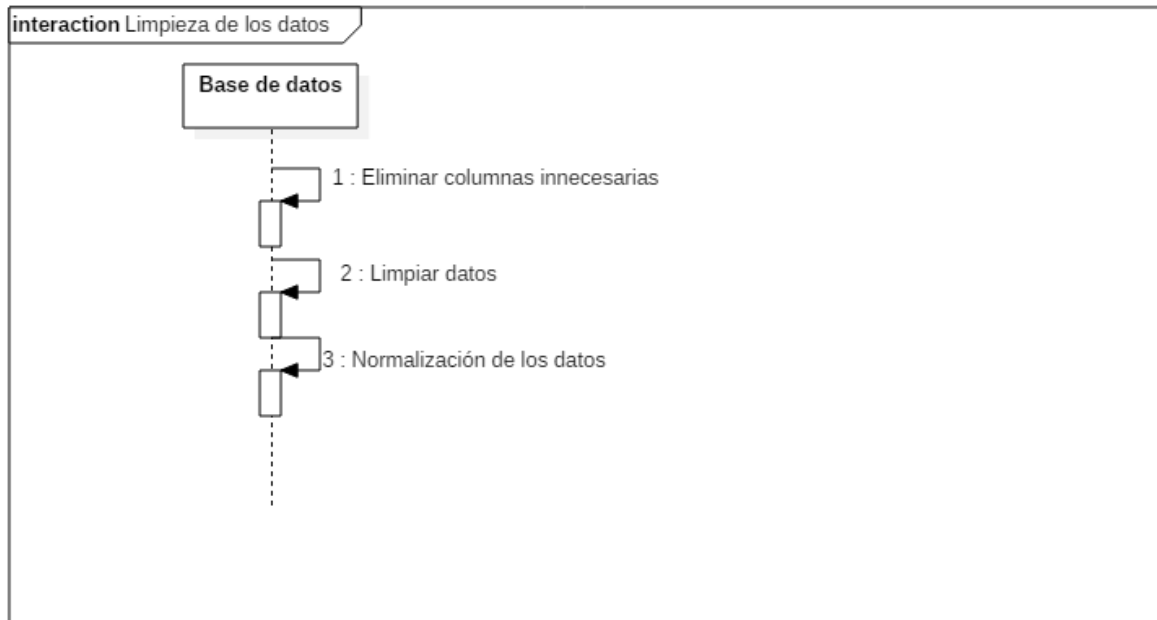


Figura 55. Diagrama de secuencia: limpieza de los datos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 39. Descripción diagrama de secuencia: limpieza de los datos

Objeto	Descripción
Base de datos	Una vez cargada la tabla, la base de datos se encargara de eliminar las columnas innecesarias, posteriormente realizara la limpieza de los datos y por último, realizara la normalización de algunos datos.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.11. Diagrama de secuencia consultar listado de pruebas

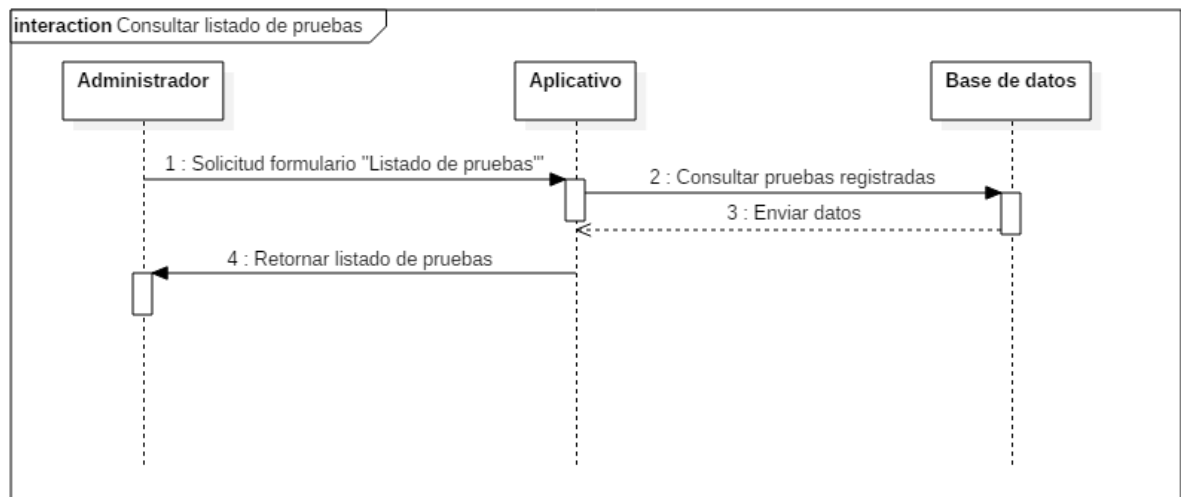


Figura 56. Diagrama de secuencia: consultar listado de pruebas

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 40. Descripción diagrama de secuencia: consultar listado de pruebas

Objeto	Descripción
Administrador	El administrador tiene una secuencia de actividades, la cual inicia con la solicitud del formulario "Listado de pruebas"
Aplicativo	El aplicativo se encargara inicialmente de solicitar a la base de datos el listado de pruebas cargadas, posteriormente se encargara de mostrarle al usuario el formulario con el listado de pruebas.
Base de datos	La base de datos se encargara de consultar el listado de pruebas registradas.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.12. Diagrama de secuencia cálculo de los indicadores de brecha de género

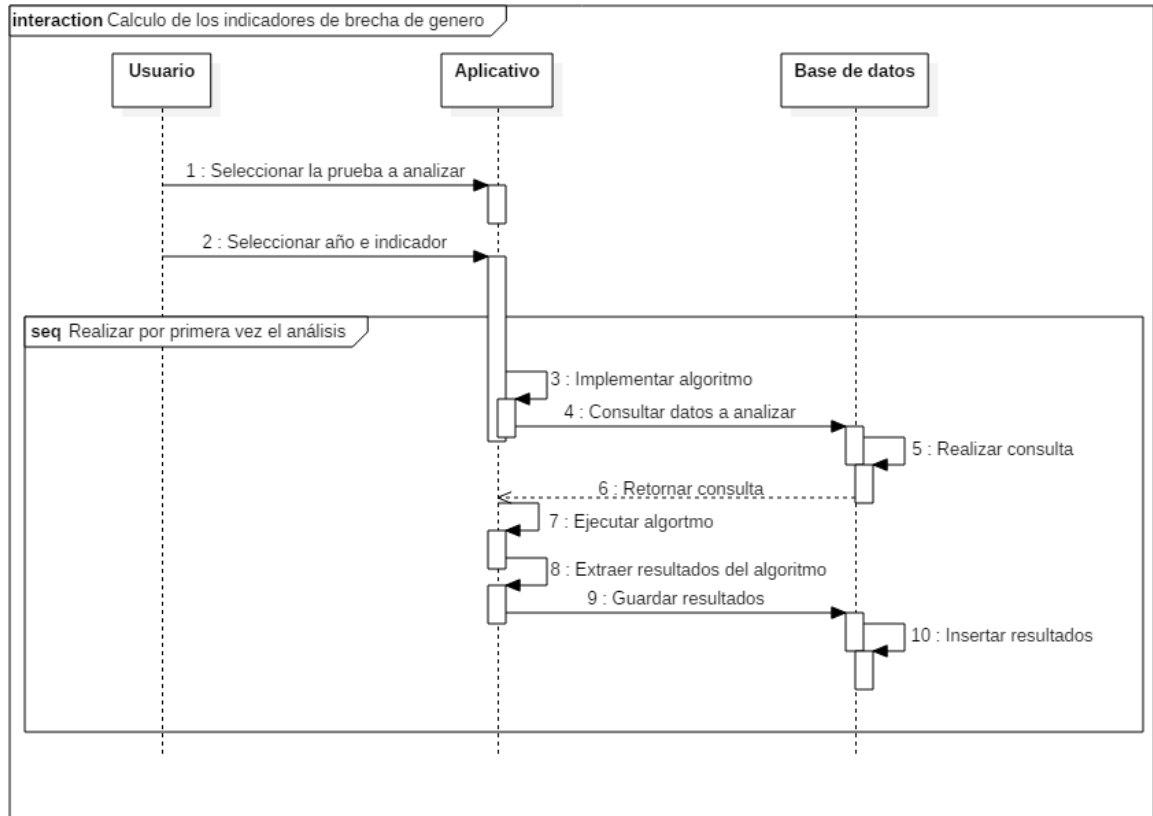


Figura 57. Diagrama de secuencia: cálculo de los indicadores de brecha de género
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 41. Descripción diagrama de secuencia: cálculo de los indicadores de brecha de género

Objeto	Descripción
Usuario	El usuario se encargara de seleccionar la prueba, el año y el indicador a analizar.
Aplicativo	El aplicativo inicialmente realizara la consulta de los datos a analizar e implementara el algoritmo K-Means. Ya con los resultados obtenidos se realizara el proceso de guardado en la base de datos.
Base de datos	La base de datos se encargara de realizar la consulta de los datos a analizar y también se encargara de registrar los resultados hallados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.13. Diagrama de secuencia análisis predictivo

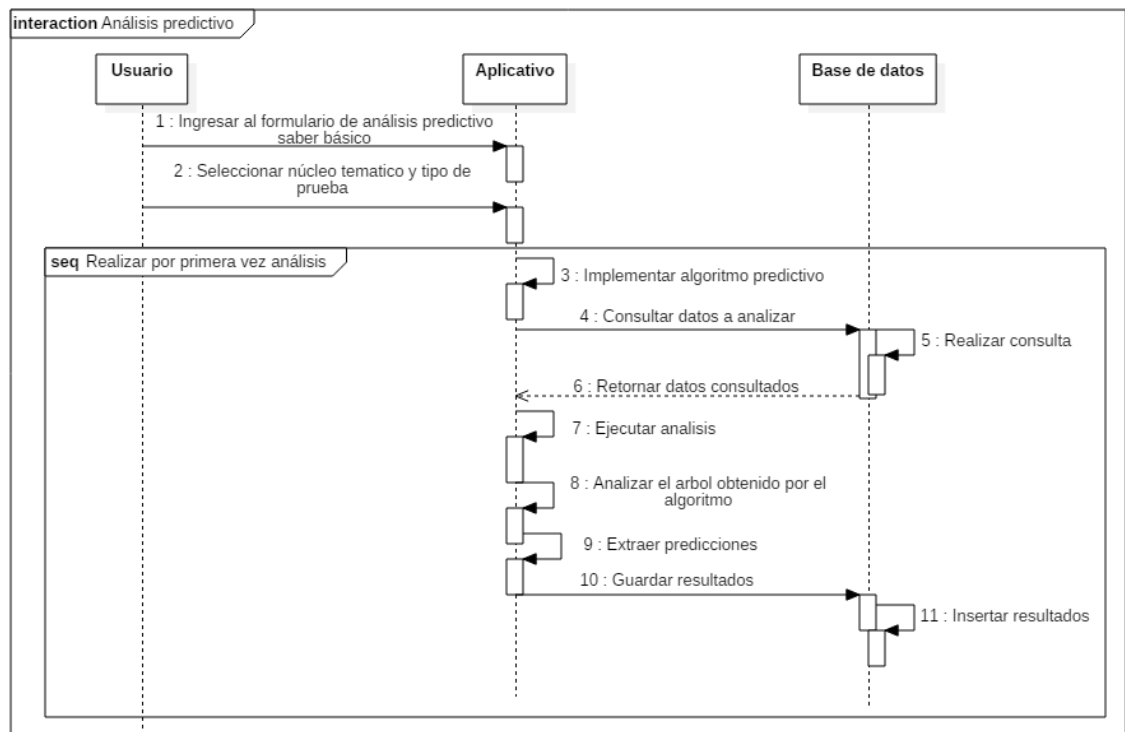


Figura 58. Diagrama de secuencia: análisis predictivo
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 42. Descripción diagrama de secuencia: análisis predictivo

Objeto	Descripción
Usuario	El usuario ingresara a la vista de análisis predictivo saber básico, allí seleccionara el núcleo temático y el tipo de prueba.
Aplicativo	El aplicativo ejecutara el algoritmo de predicción J48, para ello realizara la consulta de los datos a analizar. Con el árbol generado, se establecerán las reglas de predicción y posteriormente se guardaran dichos resultados.
Base de datos	La base de datos se encargara de realizar la consulta de los datos a analizar y también de registrar los resultados hallados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.14. Diagrama de secuencia consultar resultados de los indicadores

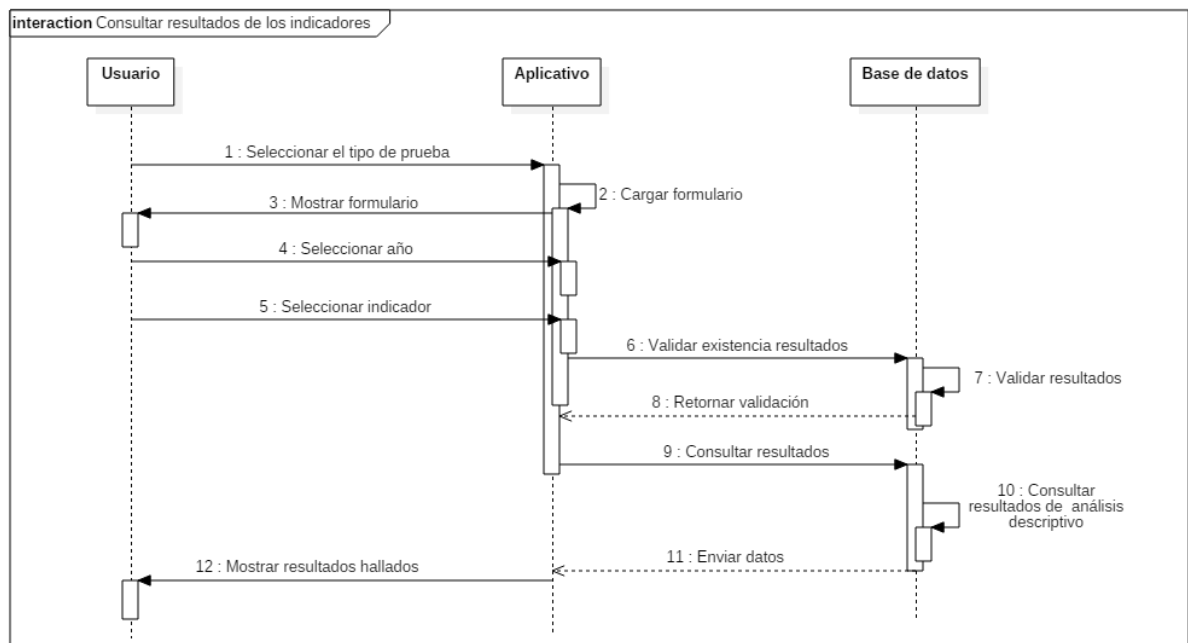


Figura 59. Diagrama de secuencia: consultar resultados de los indicadores
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 43. Descripción diagrama de secuencia: consultar resultados de los indicadores

Objeto	Descripción
Aplicativo	El aplicativo tiene una secuencia de actividades, la cual inicia con la muestra del formulario de resultados, Posteriormente realizara la consulta de los resultados registrados por indicador, año y prueba a base de datos, los cuales serán mostrados al usuario con gráficos estadísticos.
Usuario	El usuario seleccionara el tipo de prueba, el indicador y el año a analizar, posteriormente el aplicativo generara los resultados del análisis con sus respectivos gráficos.
Base de datos	La base de datos se encargara de consultar inicialmente la existencia de resultados, en caso de que existan registros, realizara la consulta de los resultados, de lo contrario procederá al cálculo de los indicadores.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.15. Diagrama de secuencia consultar resultados del análisis predictivo

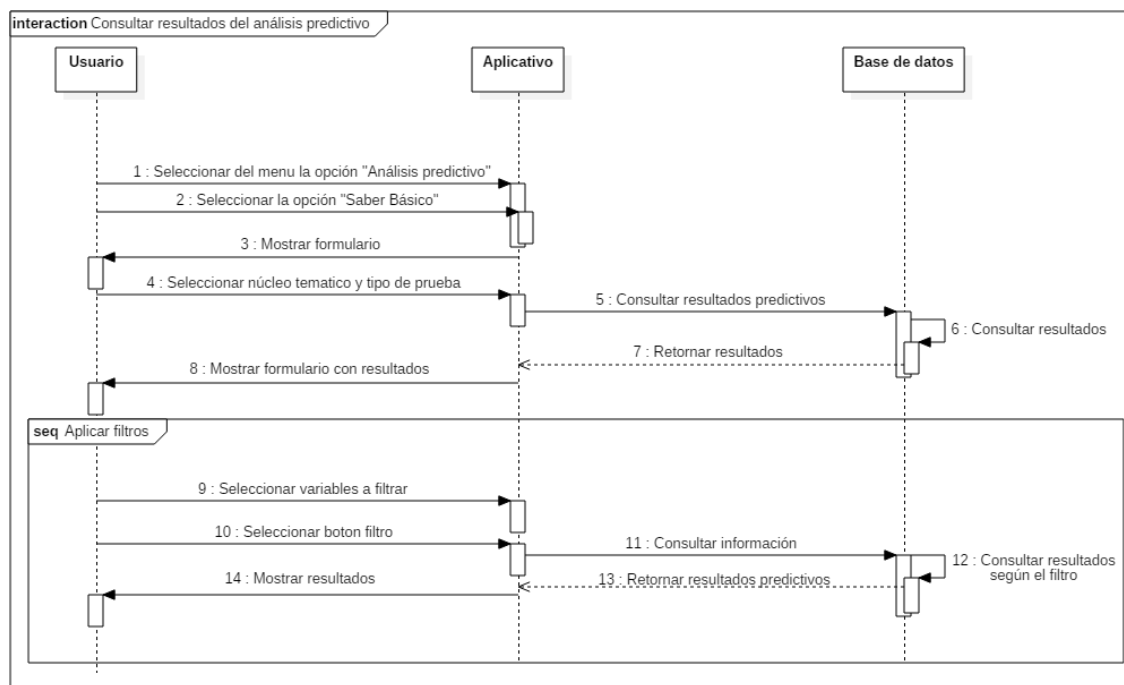


Figura 60. Diagrama de secuencia: consultar resultados del análisis predictivo
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 44. Descripción diagrama de secuencia: consultar resultados del análisis predictivo

Objeto	Descripción
Aplicativo	El aplicativo tiene una secuencia de actividades la cual inicia mostrando el formulario de análisis predictivo saber básico. Una vez el usuario seleccione el núcleo temático y el tipo de prueba, el aplicativo realizara la consulta a base de datos, si el usuario selecciona algún filtro el aplicativo se encargara de solicitar dichas consultas a base de datos. Una vez retornados los resultados, el aplicativo se encargara de generar gráficas.
Usuario	El usuario seleccionara del menú principal la opción "Análisis predictivo" posteriormente la opción "Saber básico", adicionalmente seleccionara el núcleo temático y el tipo de prueba a analizar, de esta manera el aplicativo mostrara los resultados hallados. También tendrá la posibilidad de realizar filtros sobre los resultados.
Base de datos	La base de datos se encargara de realizar las consultas de los resultados iniciales y de los filtros que solicite el usuario.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.16. Diagrama de secuencia reportes

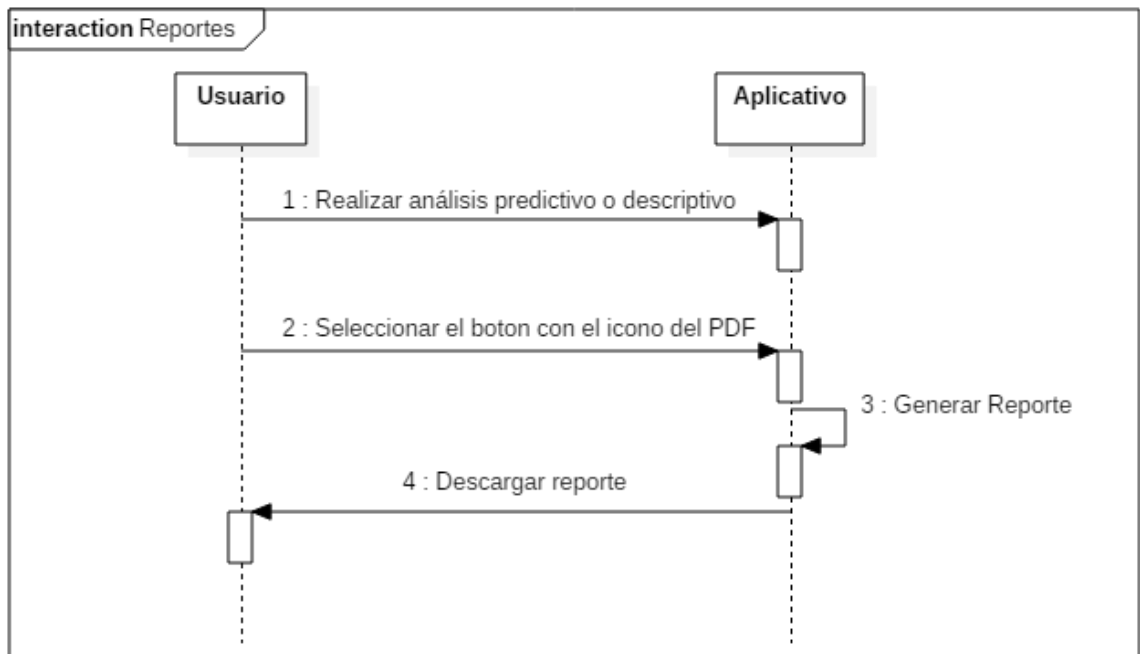


Figura 61. Diagrama de secuencia: reportes

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 45. Descripción diagrama de secuencia: reportes

Objeto	Descripción
Usuario	El usuario tiene una secuencia de actividades, la cual inicia con la generación de resultados ya sean predictivos o descriptivos, una vez allí podrá descargar los reportes en formato PDF.
Aplicativo	El aplicativo se encargara de generar reportes estadísticos en PDF, los cuales se descargarán una vez el usuario proceda a realizar el proceso.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.17. Diagrama de secuencia consultar registro de mujeres por prueba

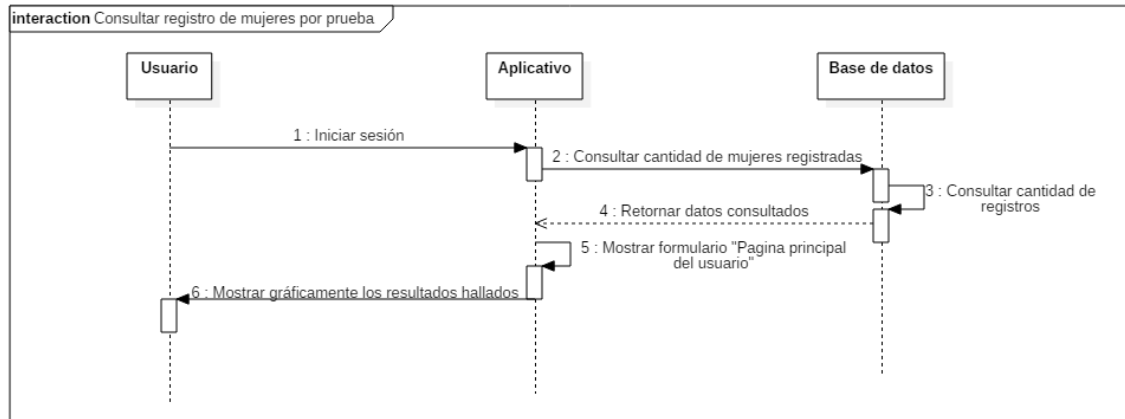


Figura 62. Diagrama de secuencia: consultar registro de mujeres por prueba
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 46. Descripción diagrama de secuencia: consultar registro de mujeres por prueba

Objeto	Descripción
Usuario	Una vez el usuario ha iniciado sesión podrá visualizar gráficamente la cantidad de mujeres registradas por prueba.
Aplicativo	El aplicativo se encargara de realizar la consulta a base de datos sobre la cantidad de mujeres registradas, una vez obtenido los resultados se le mostrara al usuario un formulario con gráficos estadísticos.
Base de datos	La base de datos se encargara de realizar la consulta de mujeres registradas por prueba.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.4.18. Diagrama de secuencia consultar cantidad de registros por prueba

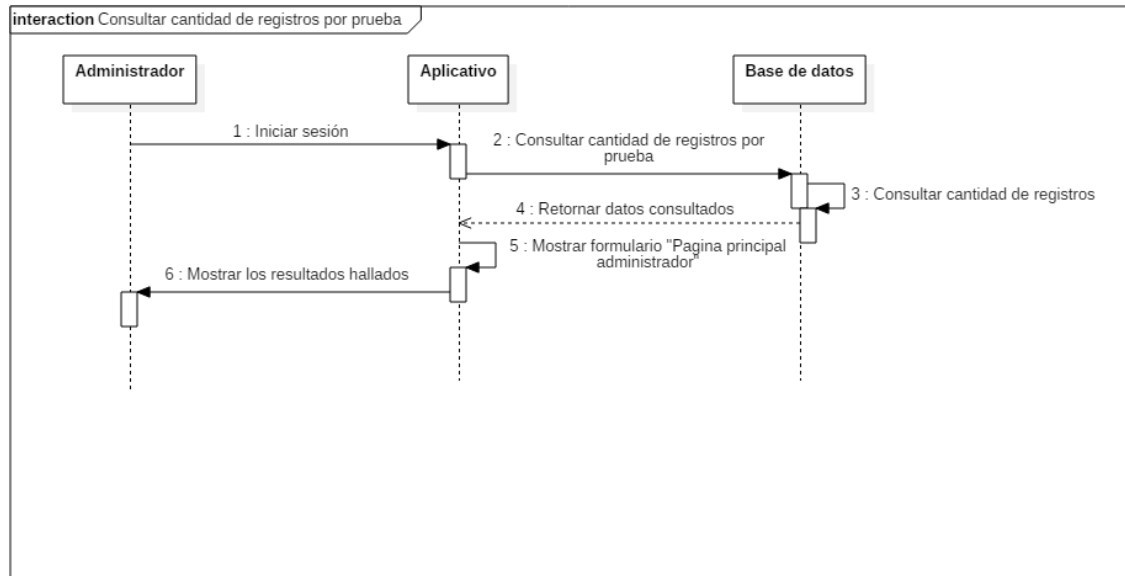


Figura 63. Diagrama de secuencia: consultar cantidad de registros por prueba
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 47. Descripción diagrama de secuencia: consultar cantidad de registros por prueba

Objeto	Descripción
Administrador	Una vez el administrador ha iniciado sesión podrá visualizar gráficamente la cantidad de registros existentes por prueba.
Aplicativo	El aplicativo se encargara de realizar la consulta a base de datos sobre la cantidad de estudiantes registrados en las pruebas, una vez obtenido los resultados se le mostrara al administrador un formulario con dicha información.
Base de datos	La base de datos se encargara de realizar la consulta de la cantidad de registros existentes por prueba.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5. Diagramas de actividad

En este diagrama se muestra un flujo de trabajo a través de una serie de acciones y roles encargados de ejecutarlas.

5.3.5.1. Diagrama de actividad inicio de sesión

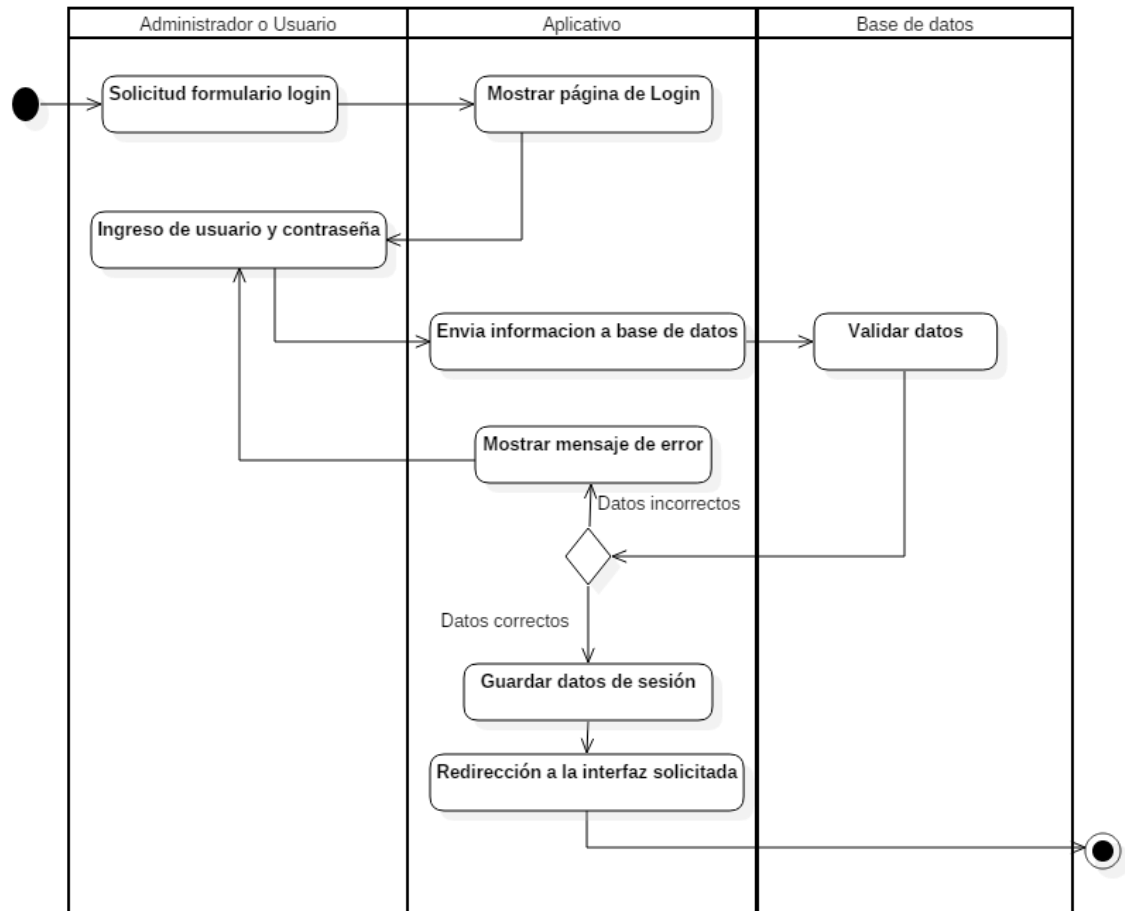


Figura 64. Diagrama de actividad: inicio de sesión
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 48. Descripción diagrama de actividad: inicio de sesión

Actividad	Descripción
Solicitud formulario Login	El usuario o administrador ingresan al aplicativo web.
Mostrar página de Login	El aplicativo procede a mostrar el formulario del Login
Ingreso de usuario y contraseña	El usuario o administrador ingresan los datos de usuario y contraseña
Enviar información a base de datos	El aplicativo envía la información ingresada por el usuario a la base de datos
Validar datos	La base de datos se encargara de validar si los datos ingresados coinciden con algún registro.
Guardar datos de sesión	El aplicativo guardara los datos de sesión del usuario o administrador que se ha logueado con anterioridad.
Redirección a la interfaz solicitada	Si los datos son correctos el aplicativo se encargara de redireccionar al módulo solicitado.
Mostrar mensaje de error	En caso de que los datos sean incorrectos el aplicativo mostrara un mensaje de error.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.2. Diagrama de actividad cerrar sesión

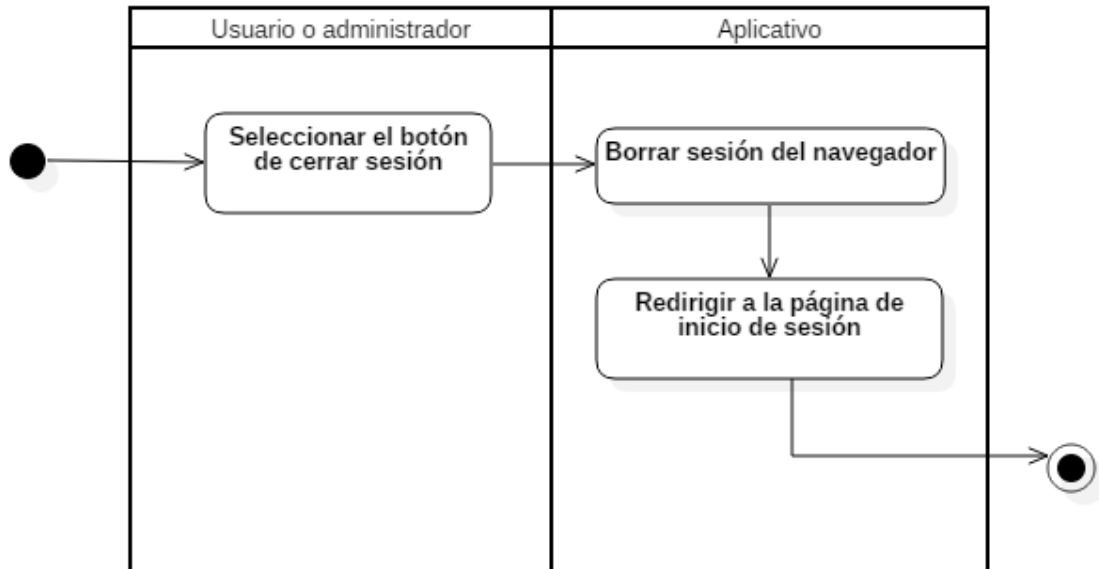


Figura 65. Diagrama de actividad: cerrar sesión
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 49. Descripción diagrama de actividad: cerrar sesión

Actividad	Descripción
Seleccionar el botón de cerrar sesión	El usuario o administrador seleccionara del menú principal el botón de cerrar sesión
Borrar sesión del navegador	El aplicativo se encargara de eliminar los datos de sesión registrados.
Redirigir a la página de inicio de sesión	El aplicativo cerrara sesión del módulo y redirigirá a la página de inicio de sesión.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

5.3.5.3. Diagrama de actividad creación usuarios

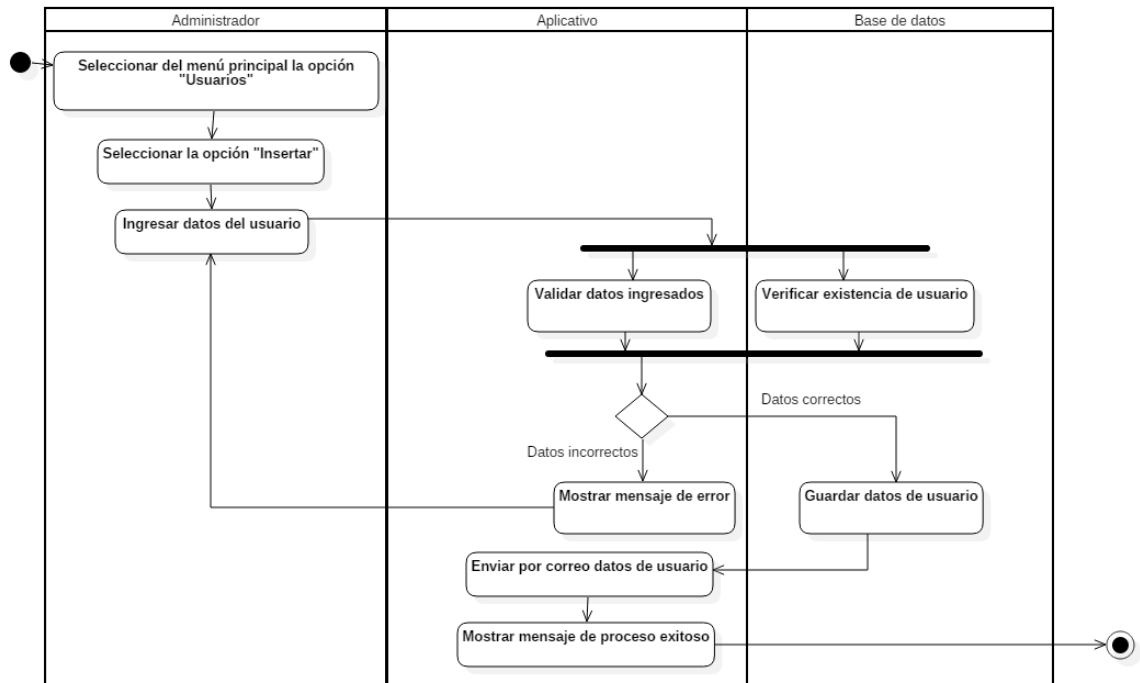


Figura 66. Diagrama de actividad: creación usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 50. Descripción diagrama de actividad: creación usuarios

Actividad	Descripción
Seleccionar del menú principal la opción "Usuarios"	Una vez iniciado sesión el administrador seleccionara del menú principal la opción "Usuarios"
Seleccionar la opción "Insertar"	El administrador seleccionara del submenú la opción "insertar", una vez seleccionado el aplicativo lo redireccionará a la página de registro de usuarios
Ingresar datos de usuario	El administrador llenara el formulario con los datos solicitados para el registro del usuario.
Validar datos ingresados	El aplicativo verifica que los datos ingresados por el administrador sean correctos y acordes a lo solicitado.
Verificar existencia de usuario	La base de datos se encargara de verificar que el usuario no este registrado.
Mostrar mensaje de error	Si los datos son incorrectos o el usuario existe el aplicativo mostrara un mensaje de error indicando la falla presentada.
Guardar datos de usuario	Si los datos son correctos y el usuario no existe la base de datos guardara la información
Enviar por correo datos de usuario	Una vez guardada la información, el aplicativo enviara un correo indicando el usuario y la contraseña asignada.
Mostrar mensaje exitoso	Posteriormente, el aplicativo mostrara un mensaje indicando que el proceso fue exitoso

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.4. Diagrama de actividad modificación usuarios

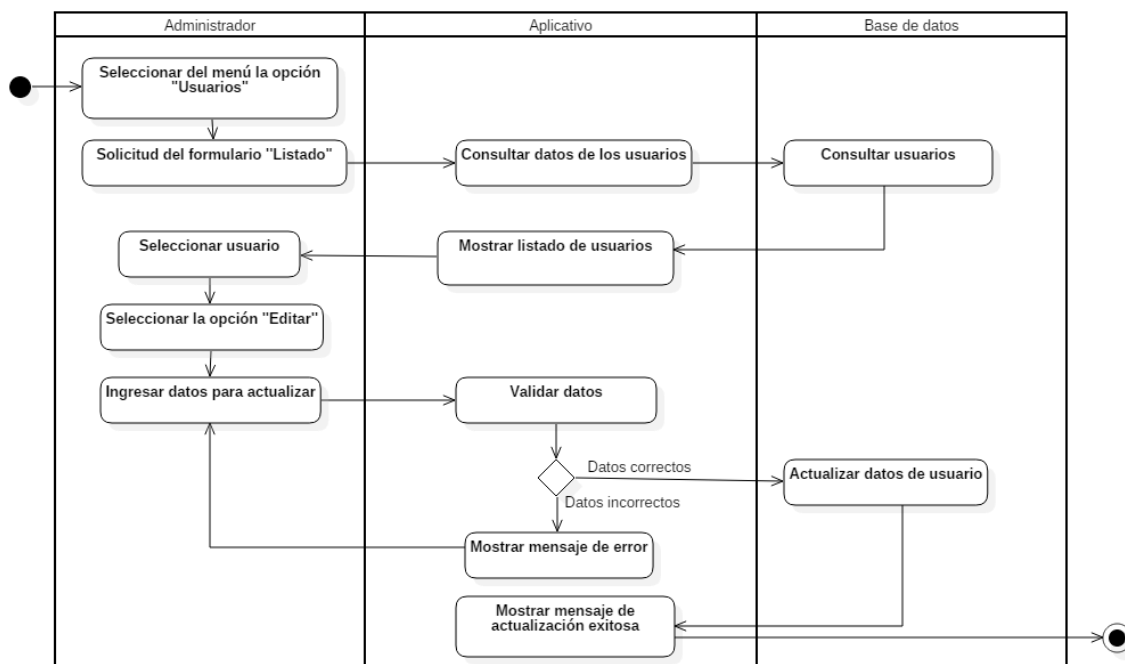


Figura 67. Diagrama de actividad: modificación usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 51. Descripción diagrama de actividad: modificación usuarios

Actividad	Descripción
Seleccionar del menú la opción "Usuarios"	El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Usuarios"
Solicitud formulario "Listado"	Dentro del submenú de usuarios seleccionara la opción de "Listado".
Consultar datos de los usuarios	El aplicativo enviara una petición de consulta sobre los usuarios registrados a la base de datos
Consultar usuarios	La base de datos se encargara de buscar los usuarios registrados en ese momento.
Mostrar listado de usuarios	El aplicativo se encargara de mostrar los resultados de la consulta en un formulario.
Seleccionar usuario	El administrador se encargara de seleccionar del listado el usuario que desea modificar.
Seleccionar opción "Editar"	Una vez seleccionado el usuario, el administrador se encargara de seleccionar la opción editar, una vez seleccionada se habilitaran las casillas para su edición.
Ingresar datos para actualizar	El administrador modificara los datos que necesita actualizar.
Validar datos	Una vez ingresada la información el aplicativo se encargara de validar que los datos estén acorde a lo solicitado.
Mostrar mensaje de error	En caso de que la información ingresada presente inconsistencias el aplicativo mostrara un mensaje de error.
Actualizar datos	Si los datos ingresados son correctos, la base de datos se encargara de actualizar la información con los nuevos datos.
Mostrar mensaje de actualización exitosa	Una vez se registre la información en la base de datos el aplicativo mostrara un mensaje indicando que la actualización fue exitosa.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.5. Diagrama de actividad consulta usuarios

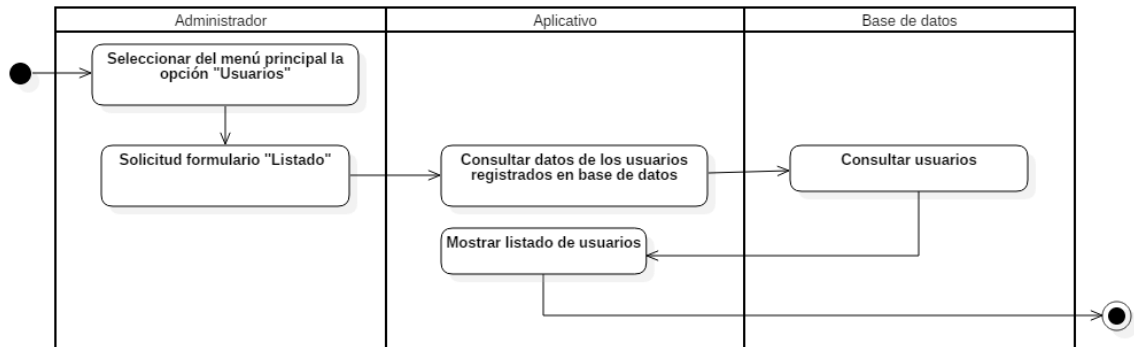


Figura 68. Diagrama de actividad: consulta usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 52. Descripción diagrama de actividad: consulta usuarios

Actividad	Descripción
Seleccionar del menú principal la opción "Usuarios"	El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Usuarios"
Solicitud formulario "Listado"	Dentro del submenú de usuarios seleccionara la opción de "Listado".
Consultar datos de los usuarios registrados en base de datos	El aplicativo enviara una petición de consulta sobre los usuarios registrados a la base de datos
Consultar usuarios	La base de datos se encargara de consultar los usuarios registrados en el momento.
Mostrar listado de usuarios	El aplicativo se encargara de mostrar los resultados de la consulta en un formulario.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.6. Diagrama de actividad eliminación usuarios

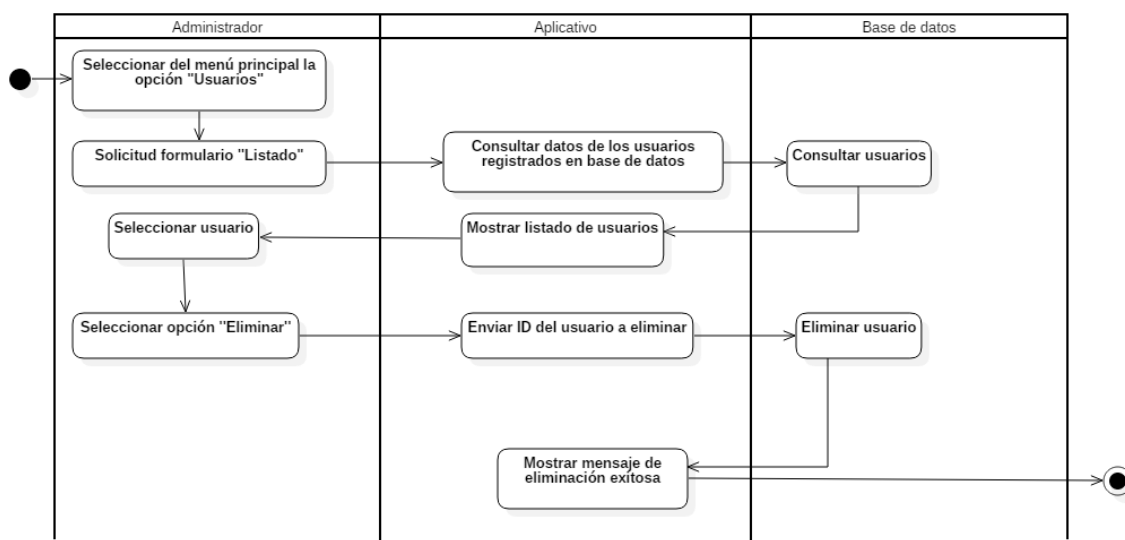


Figura 69. Diagrama de actividad: eliminación usuarios
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 53. Descripción diagrama de actividad: eliminación usuarios

Actividad	Descripción
Seleccionar del menú principal la opción "Usuarios"	El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Usuarios"
Solicitud formulario "Listado"	Dentro del submenú de usuarios seleccionara la opción de "Listado".
Consultar datos de los usuarios registrados en base de datos	El aplicativo enviara una petición de consulta sobre los usuarios registrados a la base de datos
Consultar usuarios	La base de datos se encargara de buscar los usuarios registrados.
Mostrar listado de usuarios	El aplicativo se encarga de mostrar los resultados de la consulta en un formulario.
Seleccionar usuario	El administrador seleccionara del listado, el usuario que desea eliminar.
Seleccionar opción "Eliminar"	Una vez seleccionado el usuario, el administrador seleccionara la opción de eliminar.
Enviar ID del usuario a eliminar	El aplicativo enviara a la base de datos el id del usuario a eliminar.

Eliminar usuario	La base de datos se encargara de eliminar el usuario.
Mostrar mensaje de eliminación exitosa	Una vez se elimine la información del usuario el aplicativo mostrara un mensaje indicando que la eliminación fue exitosa.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.7. Diagrama de actividad cargar pruebas

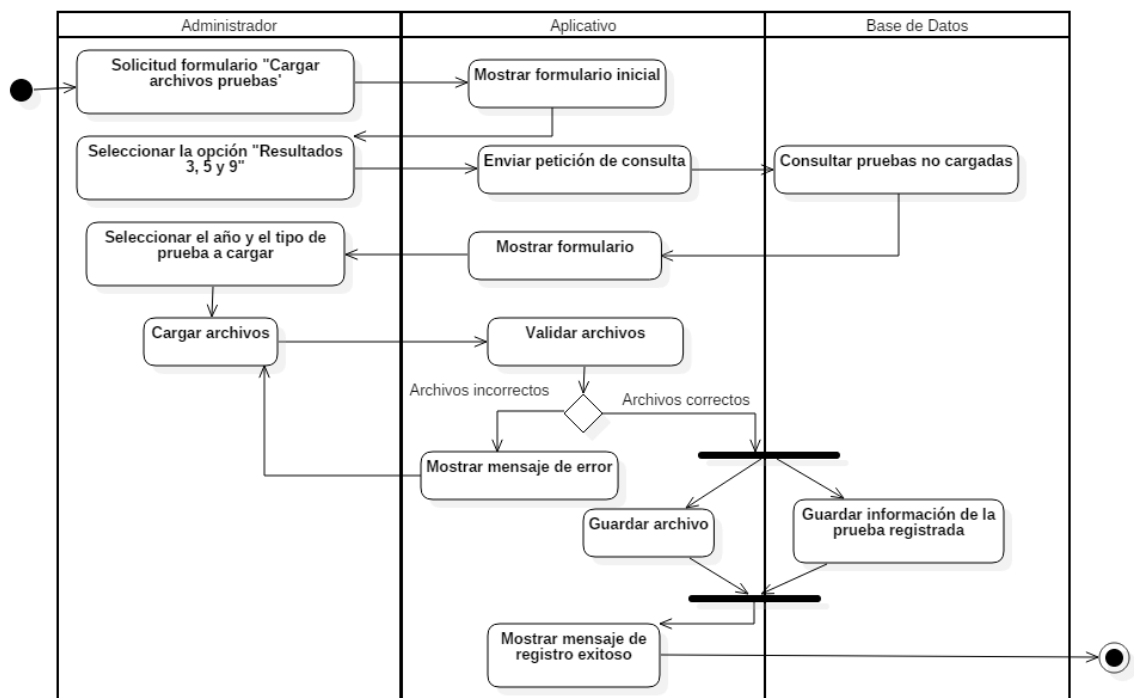


Figura 70. Diagrama de actividad: cargar pruebas
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 54. Descripción diagrama de actividad: cargar pruebas

Actividad	Descripción
Solicitud formulario "Cargar archivos pruebas"	El administrador una vez ha iniciado sesión, seleccionara del menú principal la opción "Cargar archivos pruebas"
Mostrar formulario inicial	El aplicativo mostrara un formulario para que el usuario indique que tipo de resultados desea cargar si resultados superior o básico
Seleccionar la opción "Resultados 3, 5 y 9"	El administrador seleccionara la opción "Resultados 3, 5 y 9"
Enviar petición de consulta	El aplicativo se encargara de enviar una petición de consulta sobre las pruebas no cargadas.
Consultar pruebas no cargadas	La base de datos se encargara de consultar que pruebas no han sido cargadas.
Mostrar formulario	El aplicativo cargara un formulario mostrando que pruebas no han sido cargadas por año y tipo.
Seleccionar el año y el tipo de prueba a cargar	El usuario seleccionara el año a registrar y el tipo de prueba a cargar.
Cargar archivos	Una vez ingresada la información de la prueba el usuario cargara los archivos ".zip" correspondientes.
Validar archivos	El aplicativo se encargara de validar si los archivos son correctos en cuanto a extensión, tamaño y contenido.
Mostrar mensaje de error	En el caso de que los archivos sean incorrectos el aplicativo mostrara un mensaje de error.
Guardar archivo	En el caso de que los archivos cumplan con los requerimientos serán almacenados en el servidor.
Guardar información de la prueba registrada	Una vez se ha guardado los archivos, se procederá a realizar el registro de la información de la prueba con la respectiva URL del archivo guardado.
Mostrar mensaje de registro exitoso	Una vez registrada la información el aplicativo retornara un mensaje indicando que el registro fue exitoso.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.8. Diagrama de actividad extracción de los datos

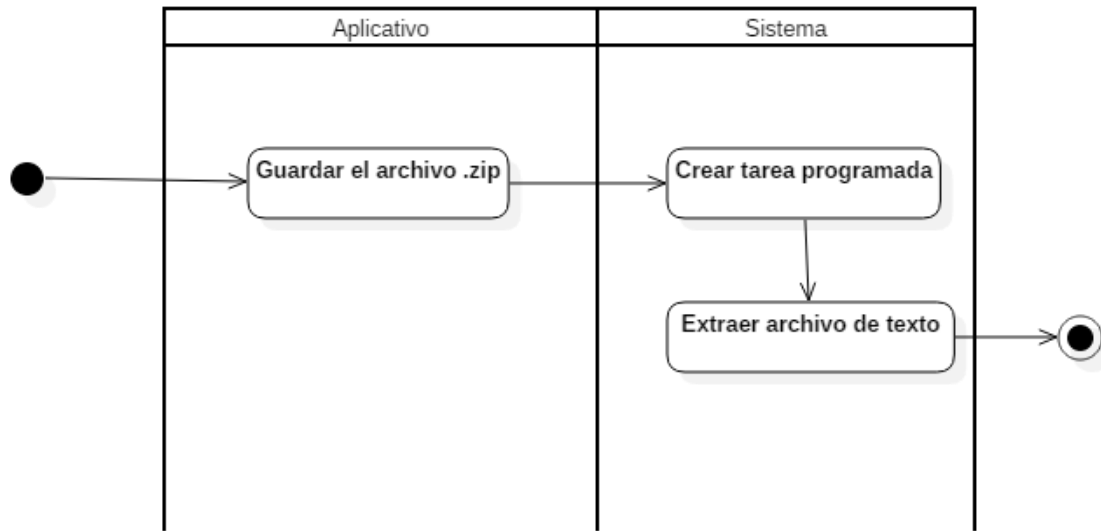


Figura 71. Diagrama de actividad: extracción de los datos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 55. Descripción diagrama de actividad: extracción de los datos

Actividad	Descripción
Guardar el archivo .zip	El aplicativo se encargara de guardar el archivo “.zip” cargado por el usuario en el servidor.
Crear tarea programada	El sistema se encargara de crear una tarea programada, la cual descomprimirá los archivos .zip y obtendrá los archivos de texto.
Extraer archivo de texto	Una vez ejecutada la tarea el sistema extraerá el archivo de texto y lo guardara.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.9. Diagrama de actividad carga de los datos

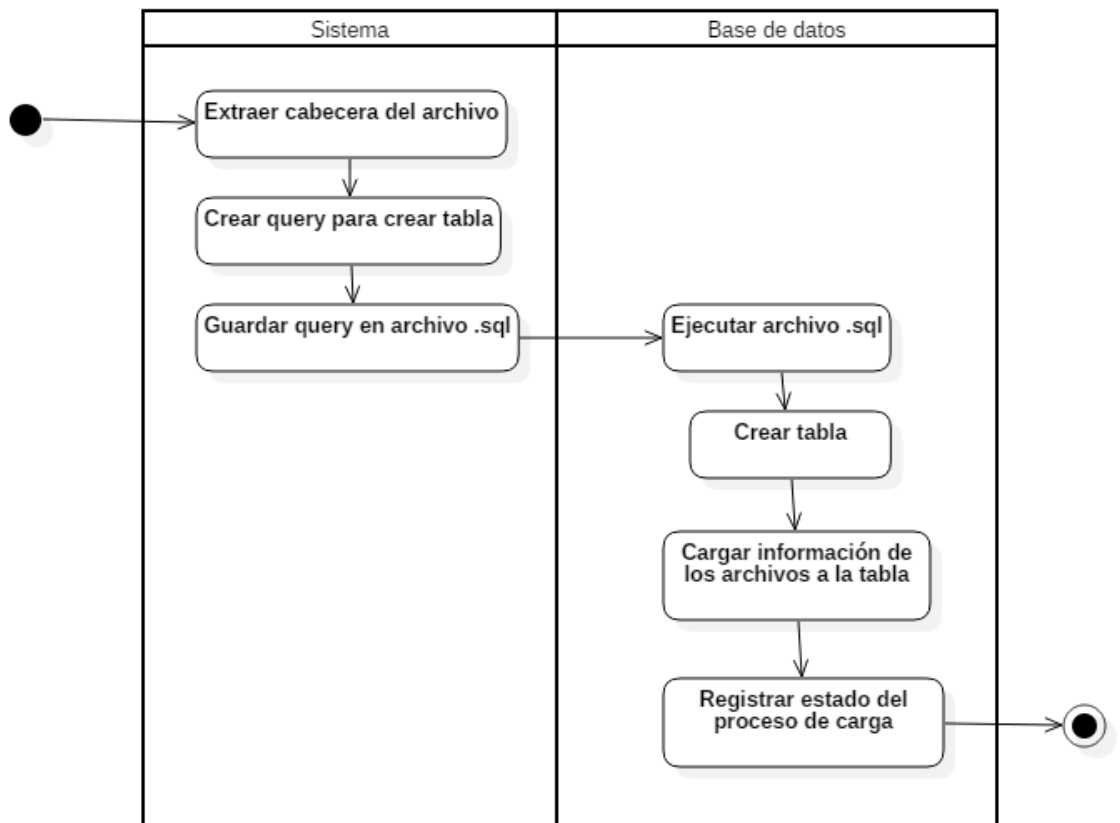


Figura 72. Diagrama de actividad: carga de los datos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 56. Descripción diagrama de actividad: carga de los datos

Actividad	Descripción
Extraer cabecera del archivo	El sistema se encargara de extraer la cabecera del archivo de texto, esto para poder establecer las columnas de la tabla.
Crear query para crear tabla	El sistema a partir de la cabecera del archivo creara un query para crear la tabla
Guardar query en archivo .sql	El sistema guardara el query en un archivo .sql
Ejecutar archivo .sql	Por medio de tareas programadas la base de datos ejecutara el archivo .sql
Crear tabla	La base de datos se encargara de crear la tabla según los datos de la cabecera y del archivo
Cargar información de los archivos a la tabla	La base de datos se encargara de cargar los datos del archivo a la tabla creada anteriormente.
Registrar estado del proceso de carga	La base de datos se encargara de registrar en la tabla "listado_tablas" el estado del proceso de carga del archivo.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.10. Diagrama de actividad limpieza de los datos

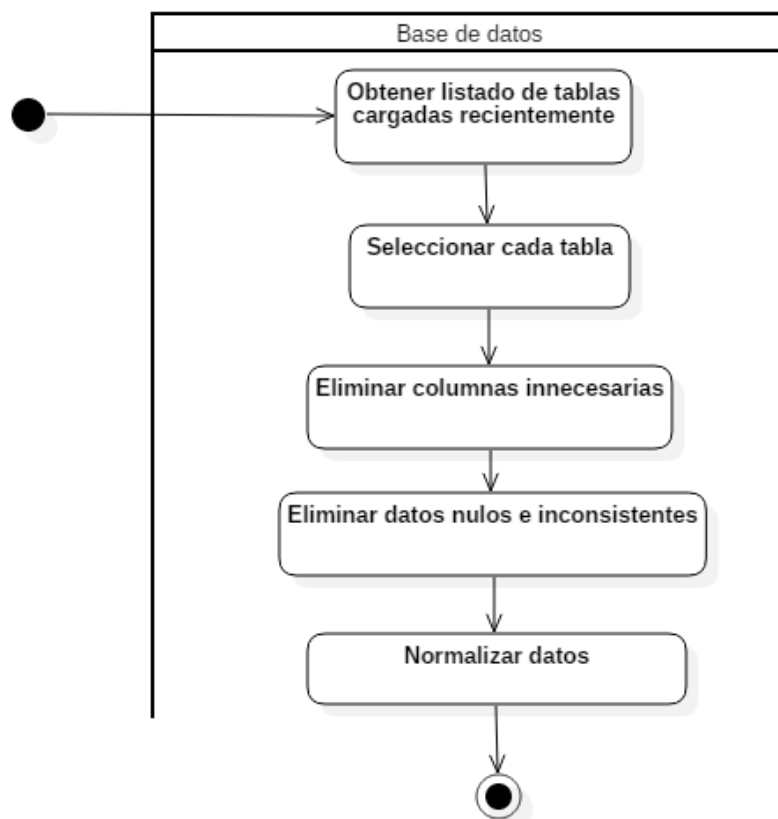


Figura 73. Diagrama de actividad: limpieza de los datos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 57. Descripción diagrama de actividad: limpieza de los datos

Actividad	Descripción
Obtener listado de tablas cargadas recientemente	La base de datos se encargara de consultar las últimas pruebas cargadas.
Seleccionar cada tabla	La base de datos seleccionara cada una de las tablas.
Eliminar columnas innecesarias	Con la tabla seleccionada la base de datos se encargara de eliminar las columnas innecesarias.
Eliminar datos nulos e inconsistentes	Posteriormente, realizara la eliminación de datos nulos o inconsistentes.
Normalizar datos	En esta actividad la base de datos se encargara de normalizar algunos campos de la tabla.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.11. Diagrama de actividad consultar listado de pruebas

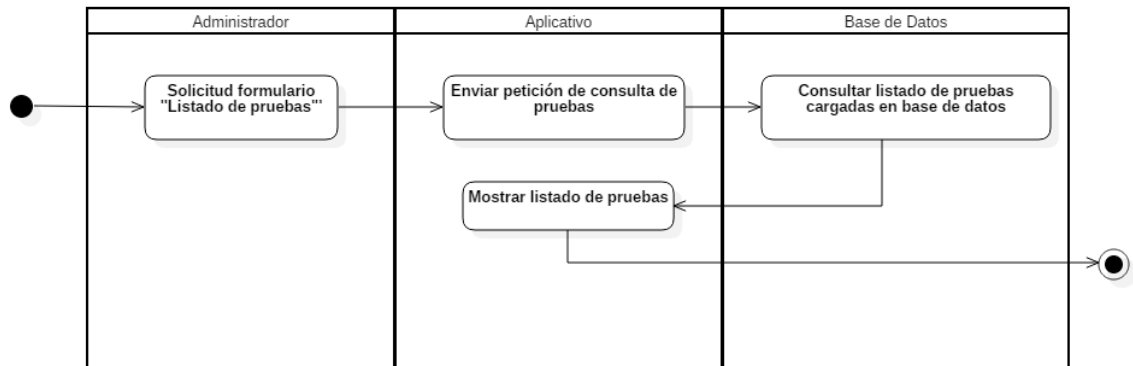


Figura 74. Diagrama de actividad: consultar listado de pruebas
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

Tabla 58. Descripción diagrama de actividad: consultar listado de pruebas

Actividad	Descripción
Solicitud formulario "Listado de pruebas"	Una vez ha iniciado sesión, el administrador seleccionara del menú principal la opción "Listado de pruebas"
Enviar petición de consulta de pruebas	El aplicativo enviara una petición de consulta a base de datos con respecto al listado de pruebas cargadas.
Consultar listado de pruebas cargadas en base de datos	La base de datos se encargara de consultar el listado de pruebas registradas.
Mostrar listado de pruebas	El aplicativo se encargara de mostrar al usuario el listado de pruebas registradas.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.12. Diagrama de actividad cálculo de los indicadores de brecha de género

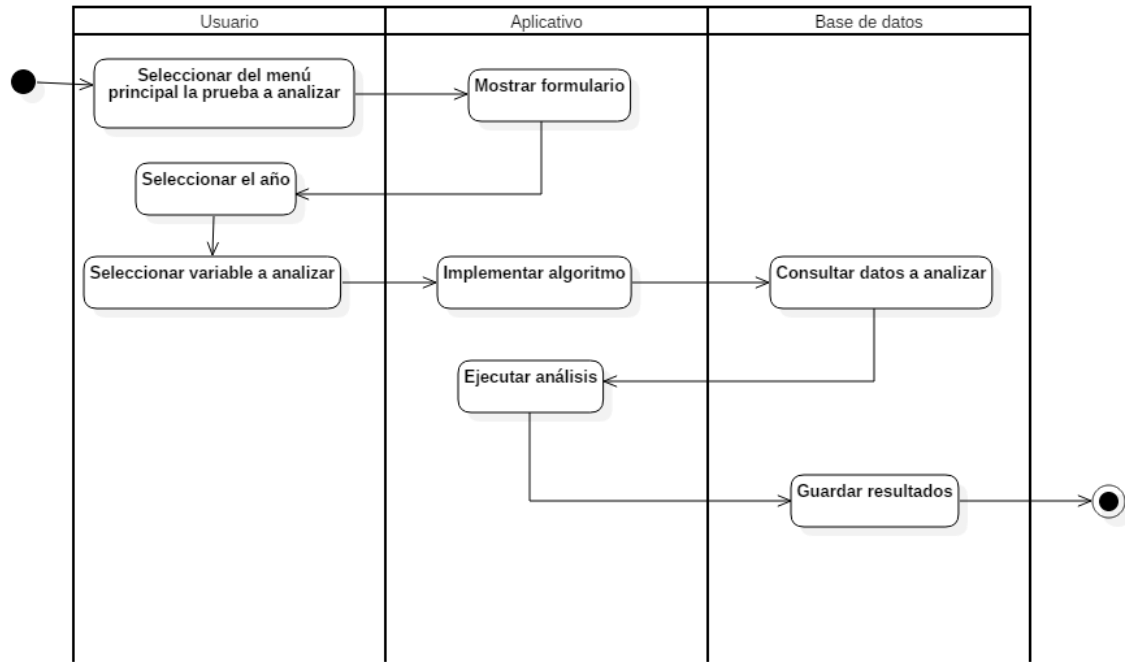


Figura 75. Diagrama de actividad: cálculo de los indicadores de brecha de género
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 59. Descripción diagrama de actividad: cálculo de los indicadores de brecha de género

Actividad	Descripción
Seleccionar del menú principal la prueba a analizar	El usuario seleccionara del menú principal el tipo de prueba a analizar
Mostrar formulario	El aplicativo mostrara un formulario, en el cual se le solicitara el año y la variable a analizar.
Seleccionar el año	El usuario seleccionara el año a analizar
Seleccionar variable a analizar	El usuario seleccionara la variable a analizar
Implementar algoritmo	El aplicativo se encargara de implementar el algoritmo K-Means de la librería de WEKA
Consultar datos a analizar	La base de datos se encargara de consultar los datos a analizar de acuerdo al tipo de prueba, el año y la variable.
Ejecutar análisis	Una vez obtenidos los datos a analizar el aplicativo ejecutara el algoritmo.
Guardar resultados	Con los resultados hallados del algoritmo la base de datos se encargara de insertarlos.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.13. Diagrama de actividad consultar resultados de los indicadores

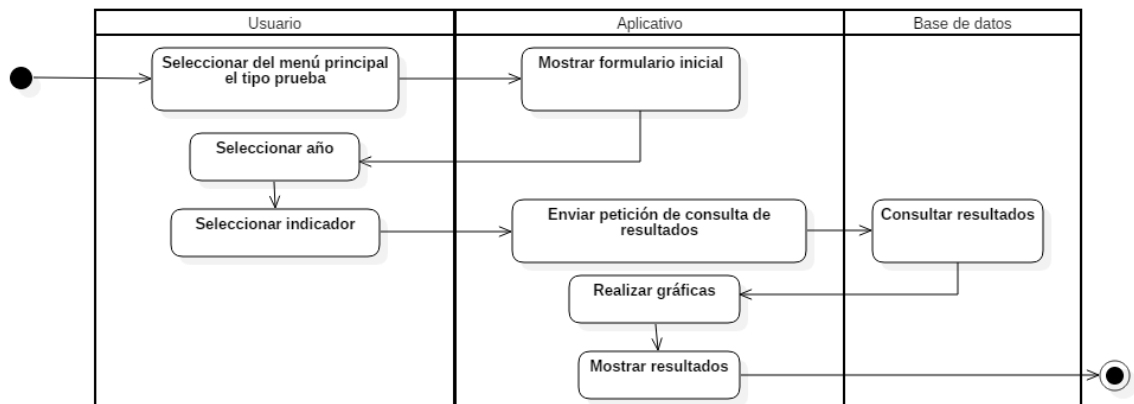


Figura 76. Diagrama de actividad: consultar resultados de los indicadores
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 60. Descripción diagrama de actividad: consultar resultados de los indicadores

Actividad	Descripción
Seleccionar del menú principal el tipo prueba	Una vez iniciada la sesión, el usuario seleccionara del menú principal la prueba a analizar
Mostrar formulario inicial	El aplicativo se encargara de mostrar un formulario en donde el usuario seleccionara el año y el indicador a analizar.
Seleccionar año	El usuario seleccionara el año a analizar.
Seleccionar indicador	El usuario seleccionara el indicador a analizar.
Enviar petición de consulta de resultados	El aplicativo se encargara de enviar una petición a la base de datos con respecto a los resultados de los indicadores.
Consultar resultados	La base de datos retornara todos los resultados relacionados con ese año, indicador y tipo de prueba.
Realizar gráficas	A partir de los resultados obtenidos, el aplicativo se encargara de generar gráficas.
Mostrar resultados	Con las gráficas generadas, el aplicativo mostrara al usuario detalladamente los resultados hallados.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.14. Diagrama de actividad análisis predictivo

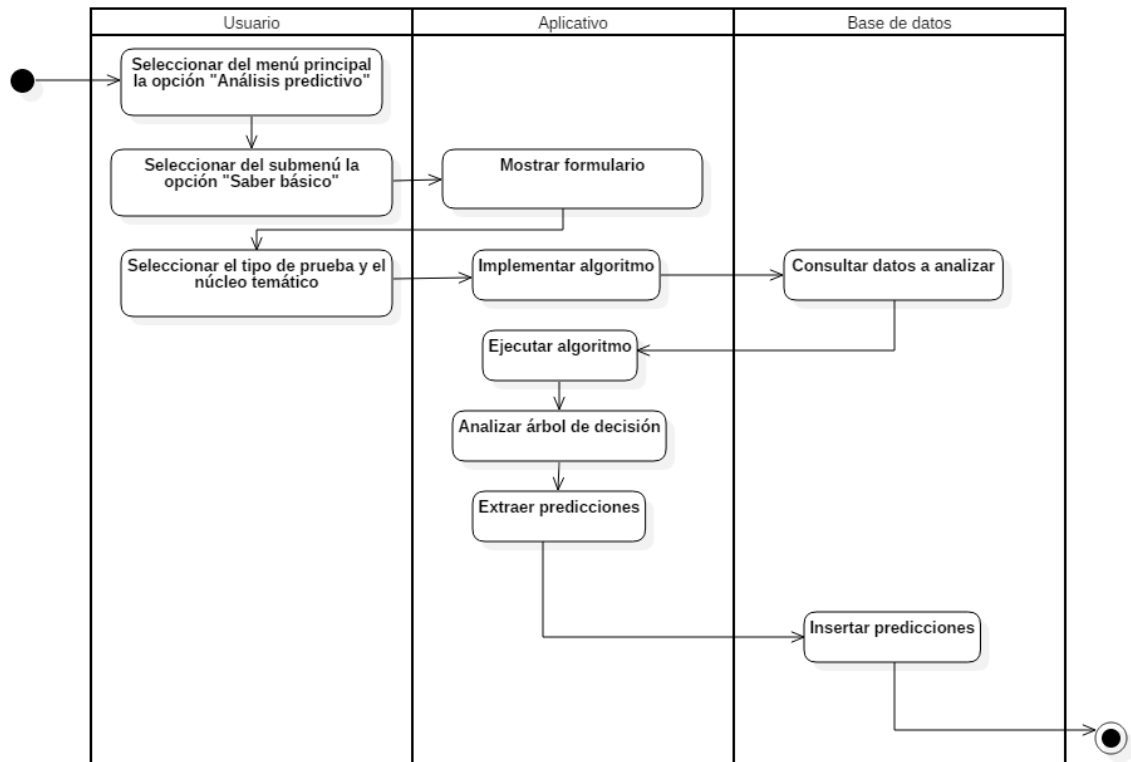


Figura 77. Diagrama de actividad: análisis predictivo
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 61. Descripción diagrama de actividad: análisis predictivo

Actividad	Descripción
Seleccionar del menú principal la opción "Análisis predictivo"	Una vez iniciada la sesión, el usuario seleccionara del menú principal la opción de "Análisis Predictivo"
Seleccionar del submenú la opción "Saber básico"	A continuación el usuario seleccionara del submenú principal la opción "Saber básico"
Mostrar formulario	El aplicativo mostrara un formulario básico donde el usuario seleccionara el tipo de prueba y el núcleo temático.
Seleccionar el tipo de prueba y el núcleo temático	El usuario seleccionara el tipo de prueba y el núcleo temático a analizar.
Implementar algoritmo	El aplicativo implementara el algoritmo J48 de la librería WEKA.
Consultar datos a analizar	La base de datos se encargara de consultar los datos a analizar, de acuerdo al tipo de prueba y al núcleo temático.
Ejecutar algoritmo	A partir de los datos consultados, el aplicativo ejecutara el algoritmo.
Analizar árbol de decisión	Con el árbol generado por el algoritmo, el aplicativo se encargara de analizarlo y establecer su matriz de confusión.
Extraer predicciones	El aplicativo extraerá las reglas de predicción del árbol de decisión.
Insertar predicciones	La base de datos se encargara de insertar los resultados predictivos.

Fuente: (Viracacha Viracacha & Barahona Rodríguez, 2020)

5.3.5.15. Diagrama de actividad consultar resultados del análisis predictivo

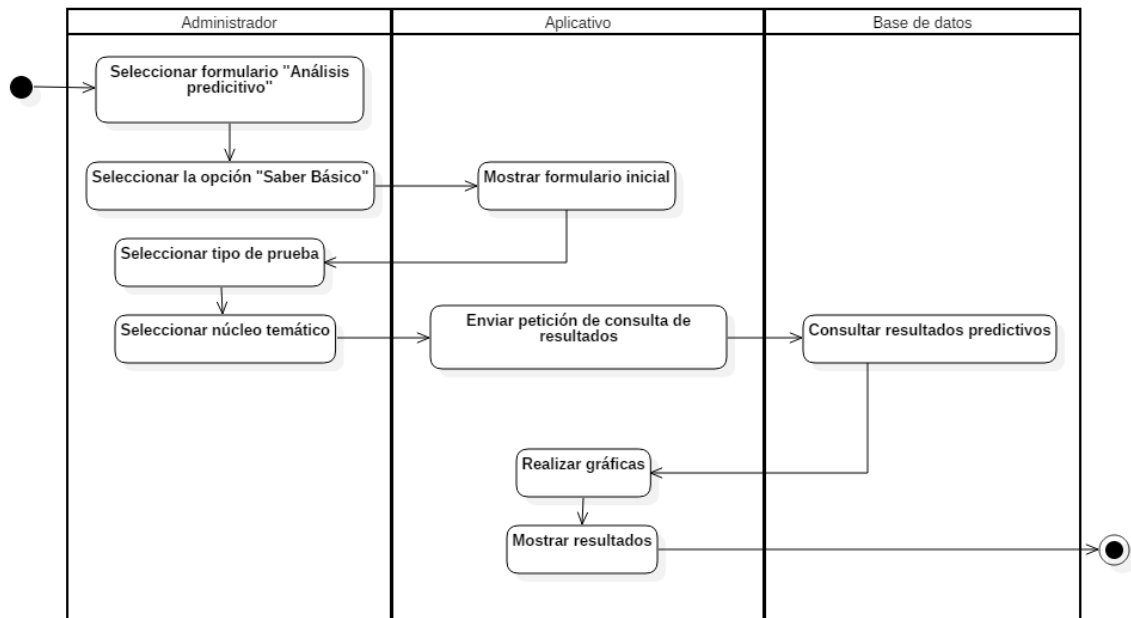


Figura 78. Diagrama de actividad: consultar resultados del análisis predictivo
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 62. Descripción diagrama de actividad: consultar resultados del análisis predictivo

Actividad	Descripción
Seleccionar del menú principal "Análisis predictivo"	Una vez iniciada sesión, el usuario seleccionara del menú principal la opción de "Análisis Predictivo".
Seleccionar del submenú la opción "Saber básico"	A continuación el usuario seleccionara del submenú principal la opción "Saber básico"
Mostrar formulario	El aplicativo mostrara un formulario básico donde el usuario seleccionara el tipo de prueba y el núcleo temático.
Seleccionar el tipo de prueba	El usuario seleccionara el tipo de prueba a analizar.
Seleccionar núcleo temático	Posteriormente a la selección del tipo de prueba el usuario seleccionara el núcleo temático a analizar
Enviar petición de consulta de resultados	El aplicativo enviara una petición de consulta a base de datos, con respecto a los resultados predictivos.
Consultar resultados predictivos	La base de datos retornara todos los resultados predictivos relacionados con el tipo de prueba y núcleo temático seleccionado.
Mostrar resultados	A partir de los resultados obtenidos, el aplicativo mostrara los resultados de manera gráfica.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.16. Diagrama de actividad reportes

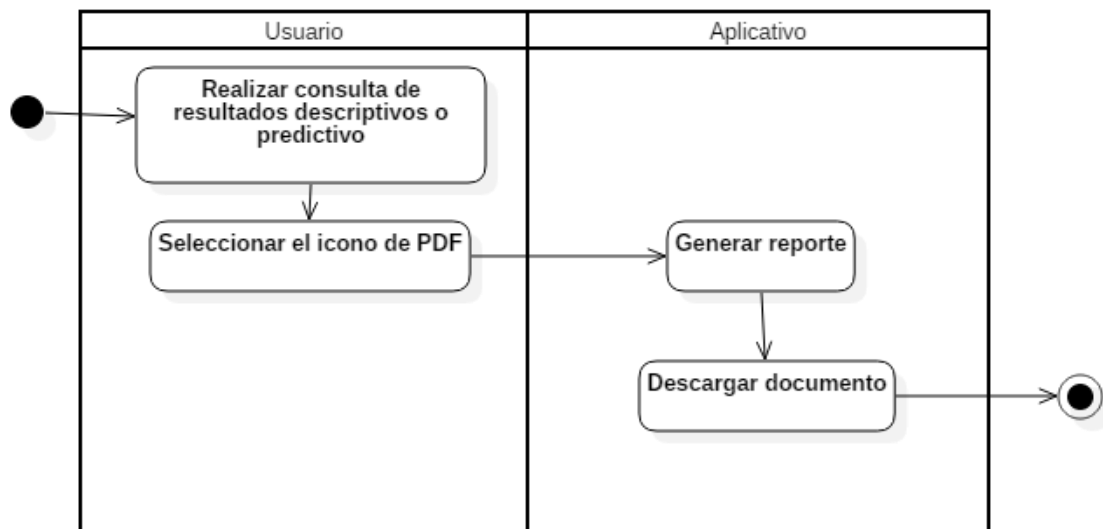


Figura 79. Diagrama de actividad: reportes
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 63. Descripción diagrama de actividad: reportes

Actividad	Descripción
Realizar consulta de resultados descriptivos o predictivo	Inicialmente se debe realizar la consulta de los resultados de los indicadores hallados o la consulta de los resultados predictivos.
Seleccionar el icono de PDF	Al final del formulario aparecerá un icono del formato PDF, el usuario lo seleccionara.
Generar reporte	Una vez seleccionado, el aplicativo creara un reporte de acuerdo al análisis previamente hecho.
Descargar documento	El aplicativo generara un reporte en PDF para que pueda ser visualizado con más detalle.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2019)

5.3.5.17. Diagrama de actividad consultar registro de mujeres por prueba

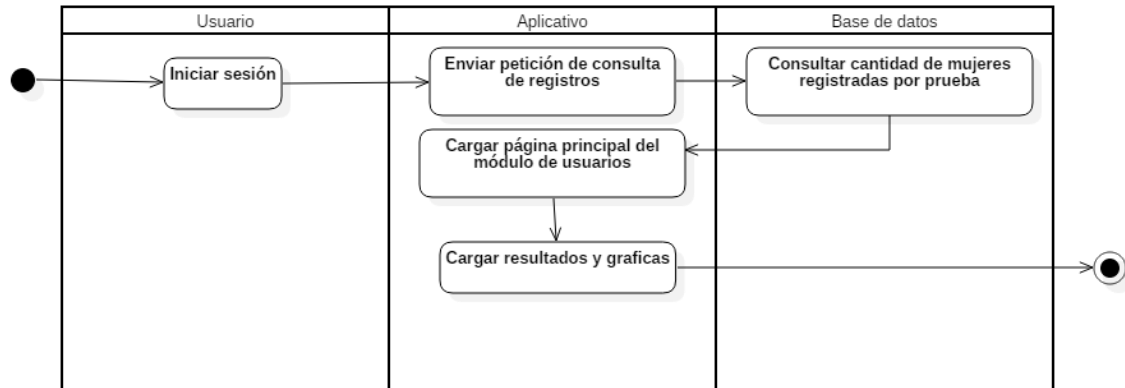


Figura 80. Diagrama de actividad: consultar registro de mujeres por prueba
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 64. Descripción diagrama de actividad: consultar registro de mujeres por prueba

Actividad	Descripción
Iniciar sesión	Inicialmente el usuario debe iniciar sesión.
Enviar petición de consulta de registros	El aplicativo enviara una petición a la base de datos, para consultar la cantidad de mujeres registradas.
Consultar cantidad de mujeres registradas por prueba	La base de datos se encargara de consultar la cantidad de mujeres registradas por prueba.
Cargar página principal del módulo de usuarios	El aplicativo se encargara de cargar la página principal del módulo de usuarios.
Cargar resultados y graficas	Con la información consultada el aplicativo se encargara de generar gráficos y mostrar los resultados al usuario.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.5.18. Diagrama de actividad: consultar cantidad de registros por prueba

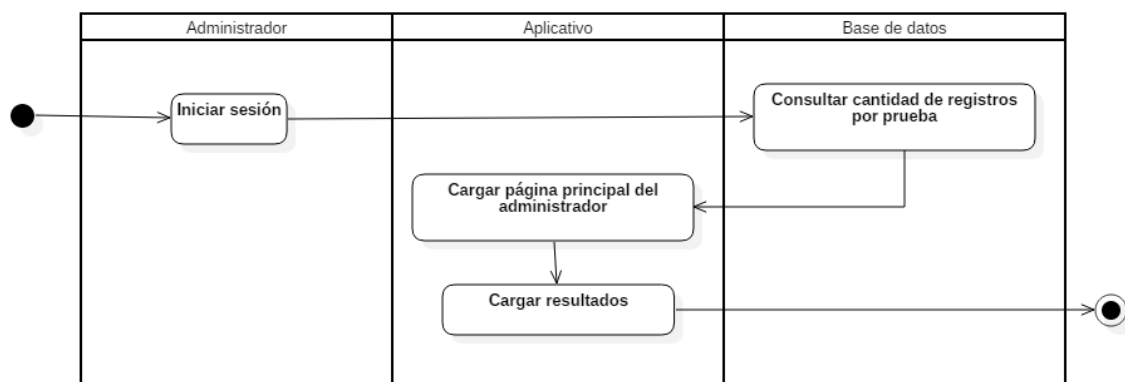


Figura 81. Diagrama de actividad: consultar cantidad de registros por prueba
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 65. Descripción diagrama de actividad: consultar cantidad de registros por prueba

Actividad	Descripción
Iniciar sesión	Inicialmente el administrador debe iniciar sesión.
Enviar petición de consulta de registros	El aplicativo enviara una petición a la base de datos, para consultar la cantidad de registros existentes.
Consultar cantidad de registros por prueba	La base de datos se encargara de consultar la cantidad de registros existentes por prueba y año.
Cargar página principal del administrador	El aplicativo se encargara de cargar la página principal del módulo administrativo.
Cargar resultados	Con la información consultada el aplicativo se encargara de mostrar los resultados al administrador.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.3.6. Diagrama de clases

Describe la estructura y el comportamiento del software de cada una de las clases y las relaciones de esta con los demás objetos.

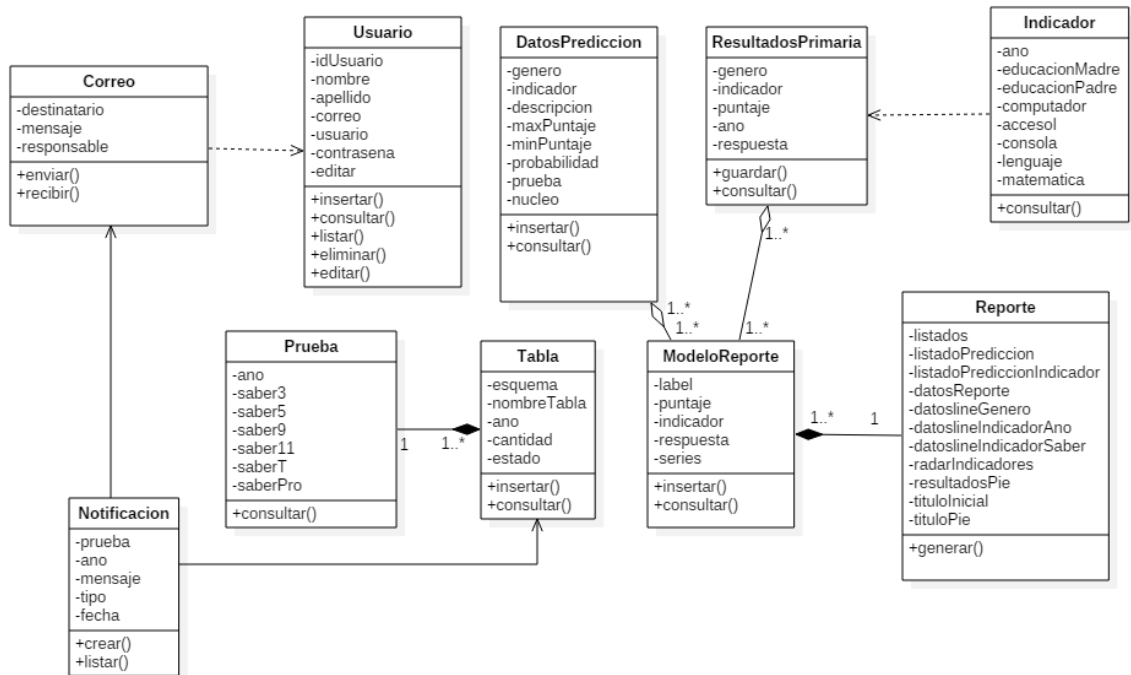


Figura 82. Diagrama de clases

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 66. Especificación de diagrama de clases

Clases	Definición
Usuario	Contiene datos de los usuarios y se encarga de registrar, consultar, modificar y eliminar datos de los usuarios.
DatosPrediccion	Una vez realizado el análisis predictivo por medio de esta clase se podrá registrar o consultar dichos resultados
ResultadosPrimaria	Una vez realizado el análisis descriptivo, esta clase se encargara de gestionar el registro o la consulta de los resultados hallados.
Indicador	Esta clase depende de ResultadosPrimaria, debido a que la clase indicador se encarga de gestionar consultas de los resultados hallados por año e indicador.
ModeloReporte	En este objeto se almacena la información que se va a presentar en el reporte
Reporte	Esta clase se encarga de conectarse con el diseñador de reportes, ya que esta envía toda la información de gráficas y datos del reporte.
Tabla	Esta clase se encarga de gestionar la inserción y la consulta de pruebas cargadas al aplicativo.
Prueba	Esta clase se encarga de realizar consultas con respecto a las pruebas que están registradas en base de datos.
Correo	Esta clase gestiona el manejo de los correos enviados (envió de correos cuando el usuario es creado) y recibidos por parte de los usuarios.
Notificacion	La clase notificación recolecta las notificaciones en cuanto a estados de proceso de carga de pruebas y en cuanto al manejo de correos entrantes.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.4. Diseños de los casos de prueba (CALISOFT)

Tabla 67. Caso de prueba: Validación de nomenclatura de base de datos

Caso de Prueba	
# Casos	1
Nombre	Validación de nomenclatura de base de datos
Requerimientos	Backup de base de datos
Propósito	Verificar que la base de datos cumpla con los estándares y la nomenclatura adecuada.
Actividades	A partir del Backup de la base de datos, el área de Calisoft se encargara de revisar que esta cumpla con los lineamientos para el nombramiento de base de datos.
Resultados	El resultado de este caso de prueba será el reporte emitido por el área.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 68. Caso de prueba: Verificación de la estructura del modelado

Caso de Prueba	
# Casos	2
Nombre	Verificación de estructura del modelado
Requerimientos	Diagrama de clases, MER, diagrama de actividades y diagrama de secuencias
Propósito	Validar la estructura y el uso adecuado de los componentes de cada diagrama
Actividades	A partir de los diagramas cargados a Calisoft, se verificara que cada uno cumpla con los estándares y el uso adecuado de los componentes de UML.
Resultados	El resultado de este caso de prueba será el reporte emitido por el área de Calisoft.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.4.1. Resultados de Calisoft

5.4.1.1. Base de datos

Componente	Total	Acertadas	Calificacion
Base de Datos	1	1	5
Esquemas	6	4	3.33
Tablas	26	17	3.27
Vistas	2	2	5
Llaves Primarias	3	3	5
Llaves Foraneas	0	0	0
Campo Descripcion	0	0	0
Campo ValorMoneda	0	0	0
Campo Observaciones	0	0	0
Obsevacion	En la mayoría de tablas se presenta el buen uso de la nomenclatura.		
Calificacion Total	4		

Figura 83. Resultados de evaluación de base de datos
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

De acuerdo a los resultados obtenidos se evidencia que en la mayoría de casos se realiza de manera correcta el nombramiento de: esquemas, base de datos, tablas, vistas, llaves y campos. Adicionalmente se observa que el 65% de las tablas y el 66% de los esquemas presentan un buen nombramiento.

5.4.1.2. Resultados diagrama de casos de uso

PROYECTO BRECHA DE GÉNERO MOD1 - Evaluación de modelado		
Diagramas de casos de uso.pdf		Diagrama de casos de uso (100%)
Evaluador	Angie Manrique	
Promedio	100%	
Componente	Validación	Observación
Actores	Correcto	
Casos de uso	Correcto	
Inclusión de casos de uso	Correcto	
Casos de uso extendidos	Correcto	
Puntos de extensión	Correcto	
Límite del sistema	Correcto	Se implementa el límite del sistema, pero se recuerda que el actor va fuera del sistema.

Figura 84. Resultados de evaluación de los diagramas de casos de uso
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto a la evaluación del diagrama de casos de uso se evidencia que cumple con todas las estructuras y nomenclaturas establecidas, dando así un resultado del 100% de correcta aplicación.

5.4.1.3. Resultados Diagramas de secuencia

Diagramas de secuencia.pdf		Diagrama de secuencias (91%)	
Evaluable	Angie Manrique		
Promedio	91%		
Componente	Validación	Observación	
Línea de vida	Correcto		
Mensajes	Correcto		
Ocurrencia de ejecución	Correcto		
Mensajes self	Correcto		
Mensajes perdidos y encontrados	Correcto		
Inicio y final de línea de vida	Correcto		
Restricciones de tiempo y duración	Incorrecto	No se presentan mensajes con restricción de tiempo, es decir, con una línea inclinada.	
Fragmentos combinados	Correcto		
Puerto	Correcto		
Descomposición en parte	Correcto		
Continuaciones / Invariantes de Estado	Correcto		

Figura 85. Resultados de evaluación de los diagramas de secuencia
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Con respecto a los resultados se observa que se cumple con la mayoría de aspectos evaluativos a excepción de las restricciones de tiempo y duración, ya que como observación se indica que dentro de los diagramas no se ve reflejado dicho aspecto.

5.4.1.4. Resultados Diagrama de actividades

Diagramas de actividades.pdf		Diagrama de actividades (86%)
Evaluador	Angie Manrique	
Promedio	86%	
Componente	Validación	Observación
Actividad	Correcto	
Acción	Correcto	
Restricciones de acción	Correcto	
Flujo de control	Correcto	
Nodo inicial	Correcto	
Nodo final de actividad	Correcto	
Nodo final de flujo	Incorrecto	El nodo final de flujo se describe como un círculo con una cruz dentro del mismo, no presenta.
Flujo de objetos	Correcto	
Nodos de decisión y combinación	Correcto	
Nodos de bifurcación y unión	Correcto	
Región de expansión	Correcto	
Gestores de excepción	Correcto	
Región de actividad interrumpible	Incorrecto	No presenta región en donde se permita la presentación de interrupciones.
Partición	Correcto	

Figura 86. Resultados de evaluación de los diagramas de actividades
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En el caso de los diagramas de actividades este presenta un 86% de buen uso de los componentes y nomenclaturas, a excepción de dos aspectos: el uso de los nodos finales de flujo y la región de actividad interrumpible.

5.4.1.5. Resultados Diagrama MER

Modelo Entidad Relación.pdf		Modelo Entidad Relación (100%)	
Evaluador	Angie Manrique		
Promedio	100%		
Componente	Validación	Observación	
Entidad	Correcto		
Atributos	Correcto		
Relación	Correcto		
Relaciones de cardinalidad	Correcto		
Claves	Correcto		

Figura 87. Resultados de evaluación del MER
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Dentro de los resultados del MER se observa que este cumple con todos los aspectos evaluativos de entidad, atributos, relaciones y claves, dando así un resultado del 100% de uso correcto.

5.4.1.6. Resultado Diagrama de clases

Diagrama de clases.pdf		Diagrama de clases (80%)	
Evaluador	Angie Manrique		
Promedio	80%		
Componente	Validación	Observación	
Interfaces	Correcto		
Tablas	Correcto		
Asociaciones	Correcto		
Generalizaciones	Incorrecto	Una generalización muestra que una clase hereda de otra clase, no se evidencia.	
Agregaciones	Correcto		
Clase asociación	Correcto		
Dependencias	Correcto		
Trazado	Incorrecto	No se evidencia trazado.	
Relaciones	Correcto	Uso correcto de relaciones entre las clases.	
Anidamientos	Correcto		
Promedio General Modelacion		91%	

Figura 88. Resultados de evaluación del diagrama de clases
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para la evaluación del diagrama de clases, se observa que cumple con la mayoría de aspectos evaluativos (80%), a excepción del uso de generalizaciones y de trazado.

5.5. Diseños de los casos de prueba (SONARQUBE)

Tabla 69. Caso de prueba: Complejidad del código

Caso de Prueba	
# Casos	1
Nombre	Complejidad del código
Requerimientos	Código fuente del aplicativo
Propósito	Determinar el nivel de complejidad de las clases, archivos y funciones
Actividades	A partir del código cargado a SonarQube, este determinara el nivel de complejidad en cuanto a funciones, clases y archivos en general.
Resultados	Nivel de complejidad

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 70. Caso de prueba: Duplicidad de código

Caso de Prueba	
# Casos	2
Nombre	Duplicidad de código
Requerimientos	Código fuente del aplicativo
Propósito	Determinar el porcentaje de duplicidad existente dentro del código.
Actividades	A partir del código cargado a SonarQube, este determinara el porcentaje de duplicidad del código.
Resultados	Nivel de duplicidad

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 71. Caso de prueba: Calidad del software

Caso de Prueba	
# Casos	2
Nombre	Calidad del software
Requerimientos	Código fuente del aplicativo
Propósito	Determinar el nivel de calidad del proyecto
Actividades	A partir del código cargado a SonarQube, este determinara el nivel de calidad del proyecto, teniendo en cuenta el estado del código.
Resultados	Nivel de calidad.

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.5.1. Resultados de SonarQube

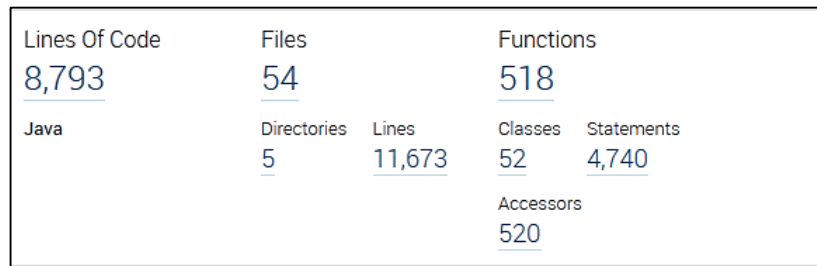


Figura 89. Métricas a evaluar del proyecto

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En total el SonarQube analizo: 8793 líneas de código, 52 clases, 518 funciones, 4740 declaraciones y 520 aspectos adicionales del código.

5.5.1.1. Complejidad del código

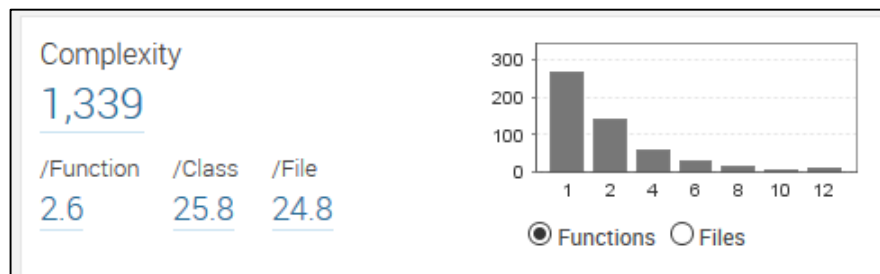


Figura 90. Complejidad del proyecto

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto a la complejidad del proyecto en general es muy baja, en el caso de las funciones esta complejidad es de 2.6 de 100, mientras que para las clases y los archivos es de 25.8 y 24.8 respectivamente, siendo estos aspectos los de mayor complejidad.

5.5.1.2. Duplicidad del código

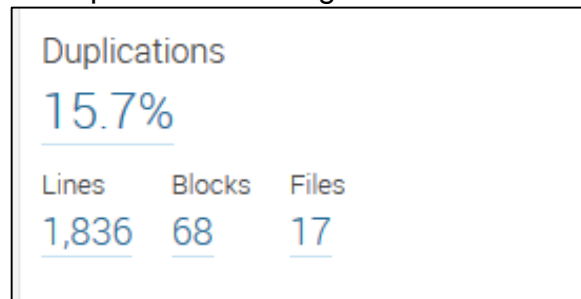


Figura 91. Duplicidad del código

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

La duplicidad es indicativo de que tan repetida esta la información, para este caso el resultado es de aproximadamente 16% demostrando que el proyecto es razonable y debe ajustarse a algunas mejoras.

5.5.1.3. Calidad del software

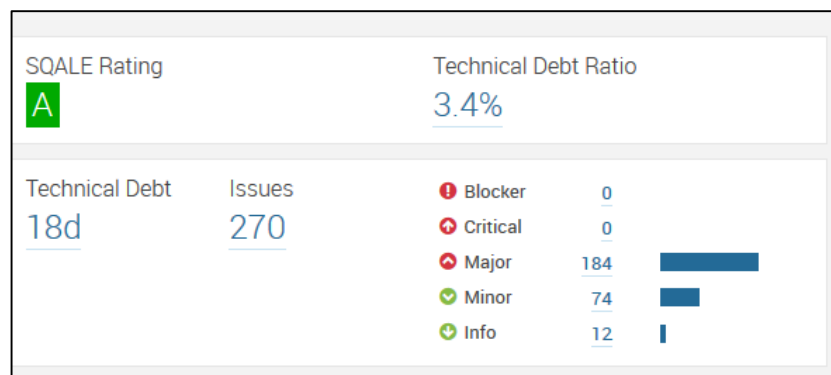


Figura 92. Calidad del software

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

De acuerdo a los resultados se observa que la calidad del proyecto es buena, el “Ratio de Deuda Técnica” indica que el proyecto es sano y no es necesario realizar muchos cambios.

5.6. Estimación de recursos

Tabla 72. Resumen de costos

Rubros	Solicitado en efectivo a UDEC	Contrapartida en especie		Total
		UDEC	Otras entidades	
Personal	\$ 35.000.000	\$ 0	\$ 0	\$ 35.000.000
Equipos	\$ 2.000.000	\$ 0	\$ 0	\$ 2.000.000
Materiales e Insumos	\$ 50.000	\$ 0	\$ 0	\$ 50.000
Servicios tecnológicos	\$ 500.000	\$ 0	\$ 0	\$ 700.000
Viajes	\$ 1.000.000	\$ 0	\$ 0	\$ 1.000.000
Otros	\$ 400.000	\$ 0	\$ 0	\$ 400.000
Totales	\$ 38.950.000	\$ 0	\$ 0	\$ 38.950.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

5.6.1. Detalles de rubros

Tabla 73. Detalle Personal

Nombre	Función en el proyecto	Tipo de vinculación	Dedicación Horas/semana	Entidad a la que pertenece	Solicitado en efectivo a UDEC	Total
Cesar Barahona	Investigador Principal	Docente	Ocho horas	Universidad de Cundinamarca	\$ 10.000.000	\$ 20.000.000
Corin Viracacha	Estudiante pregrado investigador auxiliar	Estudiante	Ocho horas	Universidad de Cundinamarca	\$ 5.000.000	\$ 5.000.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 74. Descripción de equipos

Descripción	Justificación	Cantidad	Valor Unitario	Solicitado en efectivo a UDEC	Total
Portátil Toshiba	Herramienta de trabajo	1	\$ 2.100.000	\$ 2.100.000	\$ 2.100.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 75. Descripción de materiales e insumos

Descripción	Justificación	Cantidad	Valor Unitario	Solicitado en efectivo a UDEC	Total
Papelería	Documentación	-	\$ 100.000	\$ 100.000	\$ 100.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 76. Descripción de servicios tecnológicos

Descripción	Justificación	Entidad	Valor Unitario	Solicitado en efectivo a UDEC	Total
Conexión a Internet	Fuente de investigación	Movistar	\$ 500.000	\$ 500.000	\$ 500.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 77. Descripción de viajes

Justificación	N° de Días	N° de personas	Cantidad	Valor Unitario	Solicitado en efectivo a UDEC	Total
Conferencias o ponencias	4	1	2	\$ 500.000	\$ 1.000.000	\$ 1.000.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Tabla 78. Descripción de otros

Descripción	Justificación	Entidad	Valor Unitario	Solicitado en efectivo a UDEC	Total
IP publica	Adquisición de una IP Publica para el despliegue del proyecto	Amazon	\$ 200.000	\$ 200.000	\$ 200.000

Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

6. RESULTADOS

6.1. SABER 3

6.1.1. Puntaje en lenguaje

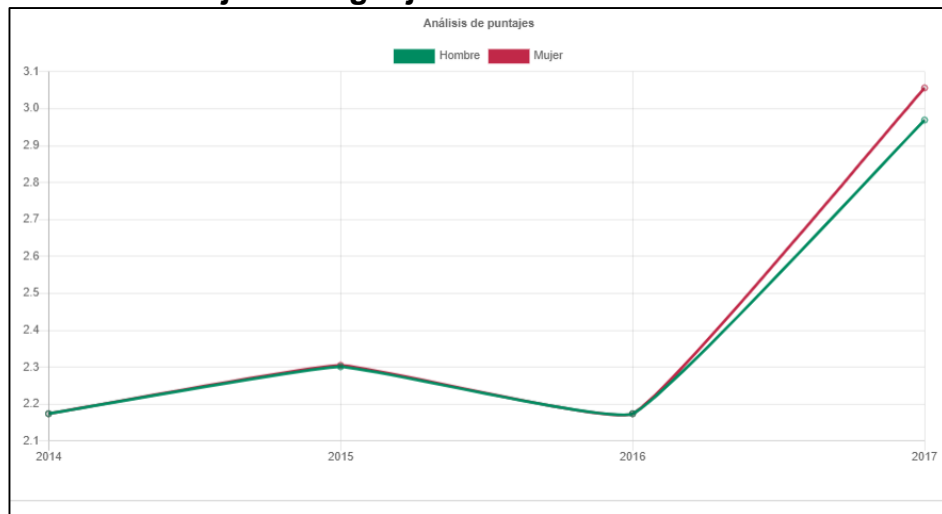


Figura 93. Resultados puntaje en lenguaje saber 3°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para los puntajes en lenguaje se observa que los resultados son equitativos hasta el año 2016, pero para el 2017 las mujeres obtienen los mejores puntajes, lo cual corrobora que en las pruebas PISA las mujeres tienden a obtener mejores resultados que los hombres en esta área.

6.1.2. Puntaje en matemáticas

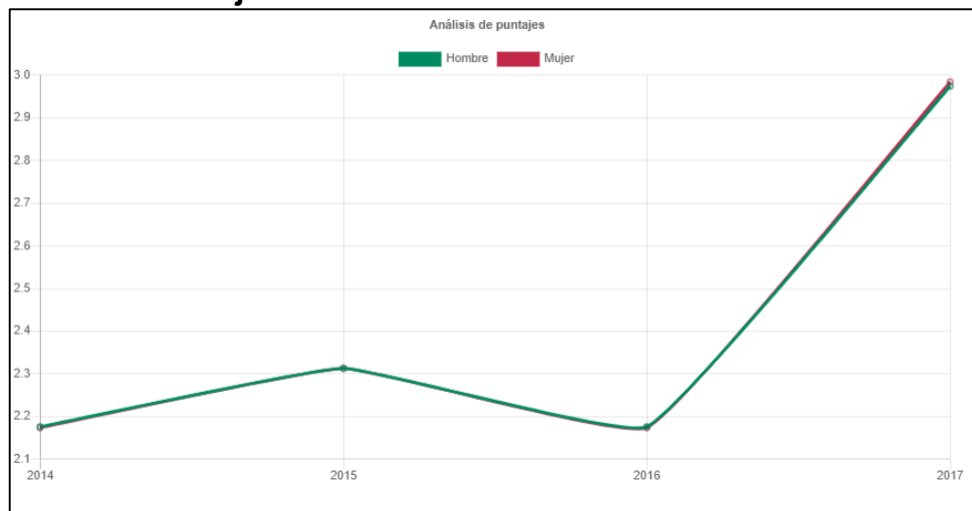


Figura 94. Resultados puntaje en matemáticas saber 3°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Con respecto a los puntajes en matemáticas se evidencia que los resultados son equitativos tanto para niñas como niños, lo cual demuestra que en esta área se ha erradicado la brecha existente en donde los hombres obtienen mejores resultados.

6.2. SABER 5

6.2.1. Educación padre

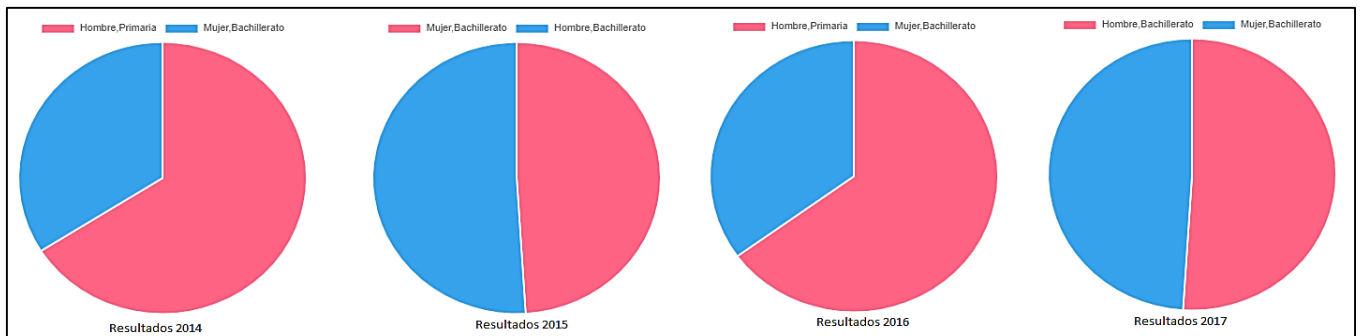


Figura 95. Resultados educación del padre en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En este indicador se puede evidenciar que existe una brecha en donde la educación de los padres de los niños para los años 2014 y 2016 logra alcanzar su primaria en un 34% en promedio.

6.2.2. Educación madre

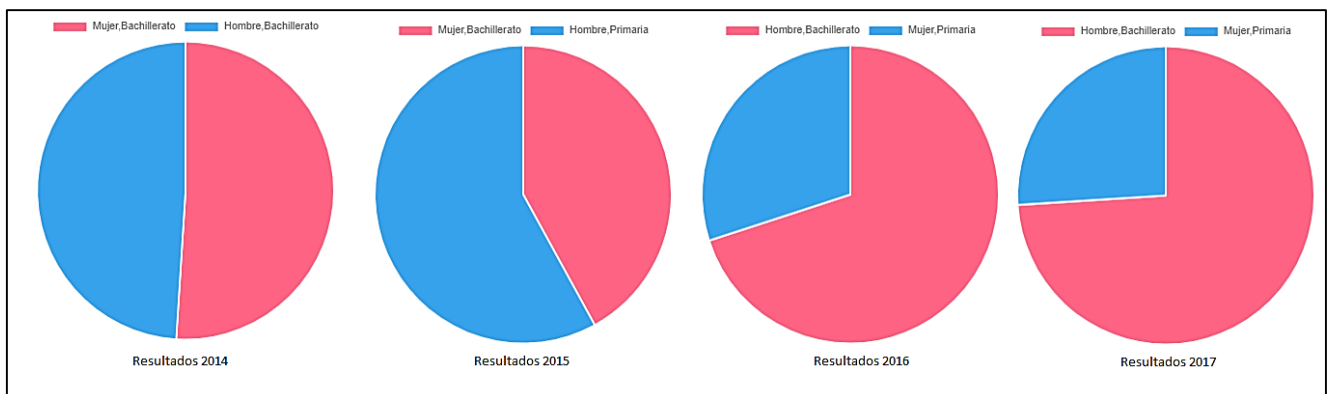


Figura 96. Resultados educación de la madre en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En el caso de la educación de la madre se evidencia que para el año 2014 es equilibrado, pero para el año 2015 las madres de los niños logran terminar su primaria en un 42%, ya para el 2016 y 2017 las madres de las niñas logran terminar su primaria en un 28% en promedio, mientras que para los niños el nivel más alto de formación es bachiller en un 68% respectivamente.

En conclusión aún se evidencia una brecha existente entre la educación de los padres y madres, la cual es a favor de ellas ya que en la mayoría de casos el mayor nivel de escolaridad alcanzado ha sido primaria.

6.2.3. Acceso a computador

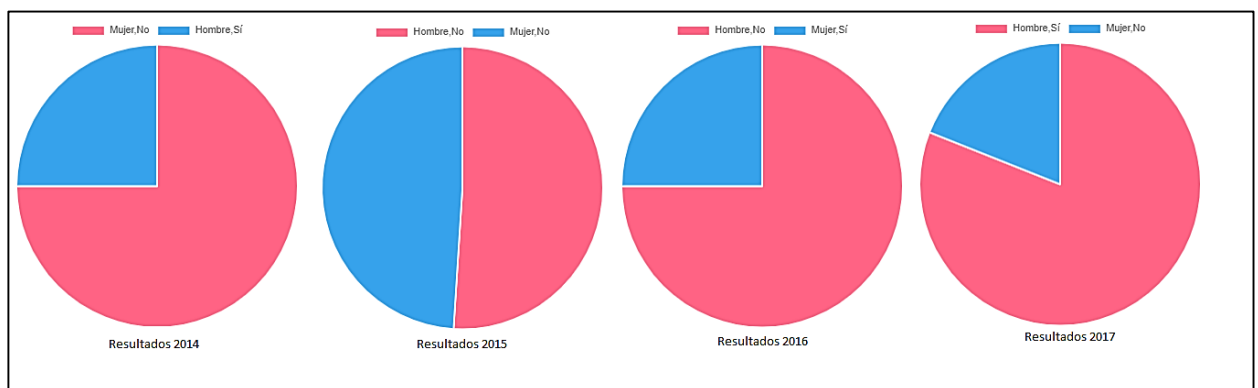


Figura 97. Resultados acceso a computador en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para los resultados de acceso al computador se evidencia que para el año 2014 y 2017 las niñas dicen no tener computador, mientras que los niños si tienen, esto corrobora la brecha tecnológica que existe desde edades tempranas. En el caso del año 2015 los resultados son equitativos. Mientras que para el 2016 sucede todo lo contrario las niñas tienen acceso al computador, en cambio los niños dicen no tener acceso a el.

6.2.4. Acceso a internet

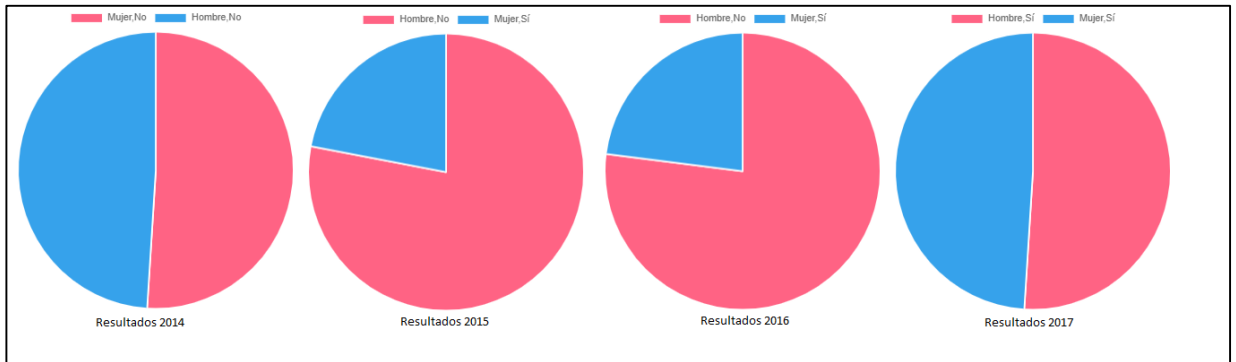


Figura 98. Resultados acceso a internet en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto al acceso al internet se evidencia que para los años 2014 y 2017 el resultado es equitativo, mientras que para el 2015 y 2016, los niños no tienen acceso en un 23 % en cambio, las niñas tienen acceso en un 77% en promedio.

6.2.5. Acceso a consola de videojuegos

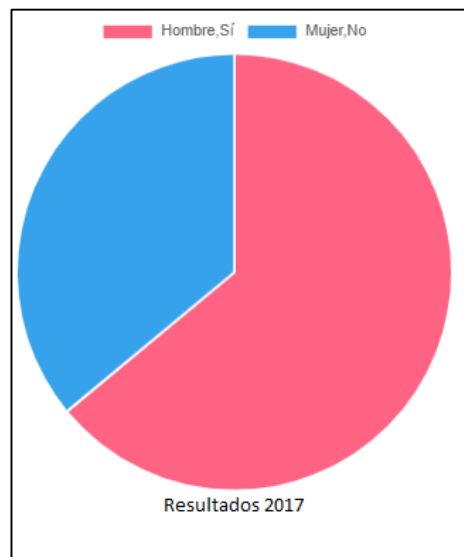


Figura 99. Resultados acceso a consola de videojuegos en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Desde el año 2017, el ICFES incluyó en sus resultados la variable acceso a la consola de videojuegos. De acuerdo a los resultados se evidencia que las niñas en

un 36% no tienen acceso, en cambio los niños tienen acceso en un 64% respectivamente, demostrando así la brecha existente en el acceso a ciertas tecnologías.

6.2.6. Resultados puntaje en lenguaje

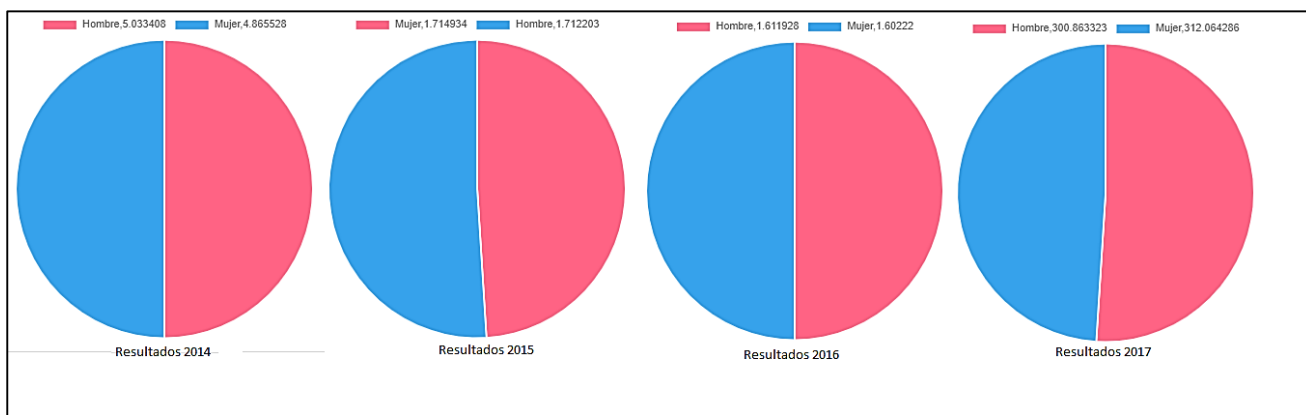


Figura 100. Resultados puntaje en lenguaje en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Cabe mencionar que el rango de puntajes para los años 2014, 2015 y 2016 es de 1 a 10 respectivamente, mientras que para el año 2017 los resultados van de 0 a 500 en general

Con respecto a los puntajes obtenidos en lenguaje para el año 2014 los niños obtienen un puntaje mayor que las niñas (1 punto de diferencia), en el caso del 2015 y 2016 los resultados son equitativos para ambos géneros y para el 2017 las mujeres obtienen 12 puntos de más que los hombres. Con esto se corrobora que para el último año evaluado las niñas tienden a obtener mejores resultados en esta área.

6.2.7. Resultados puntaje en matemáticas

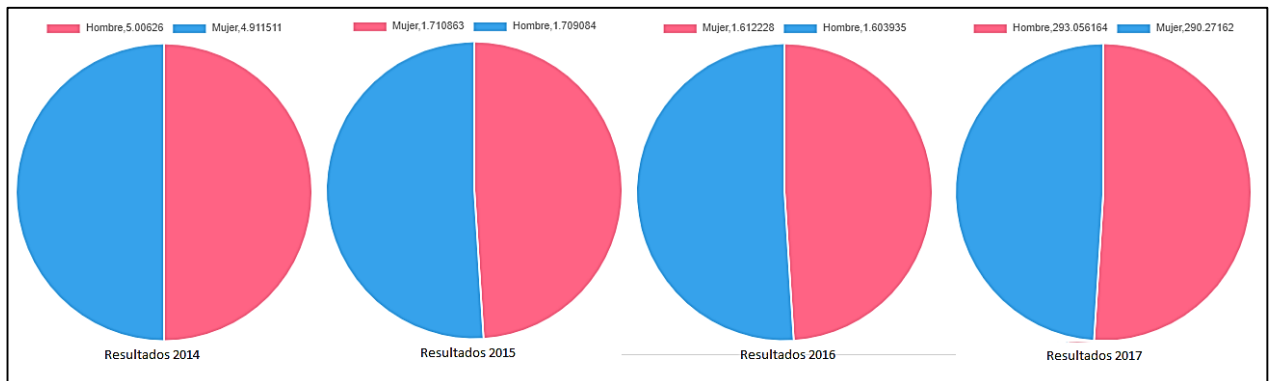


Figura 101. Resultados puntaje en matemáticas en pruebas saber 5°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En este caso se observa que para el año 2014 y 2017, los niños obtienen mejores resultados diferenciados por uno y tres puntos respectivamente frente a las niñas en el área de matemáticas, mientras que para el 2015 y 2016 los resultados se diferencian por un punto a favor de las mujeres. Demostrando así que los mejores resultados los obtienen los hombres, tal como se evidencia en las pruebas PISA.

6.3. SABER 9

6.3.1. Educación del padre

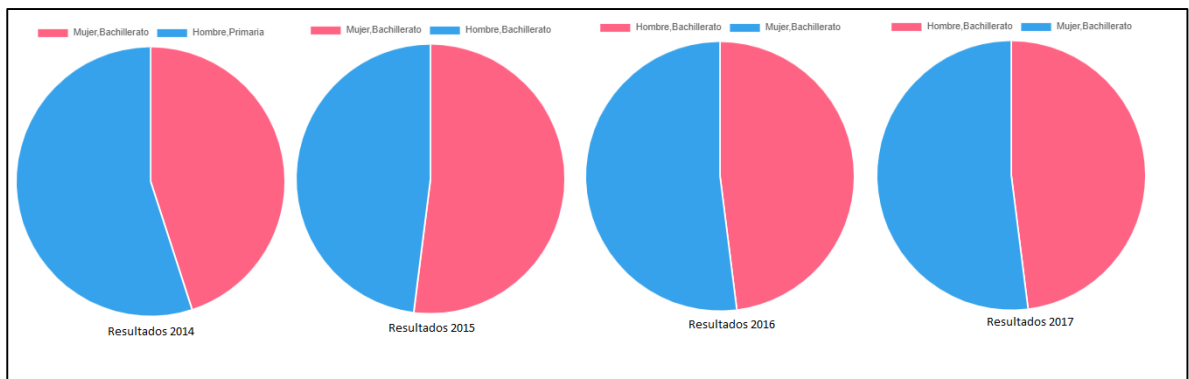


Figura 102. Resultados educación del padre en pruebas saber 9°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para este indicador los resultados son equitativos a excepción del año 2014, en donde, los resultados para los niños indica que los padres solo logran hasta primaria y los padres de las niñas logran el bachillerato.

6.3.2. Educación de la madre

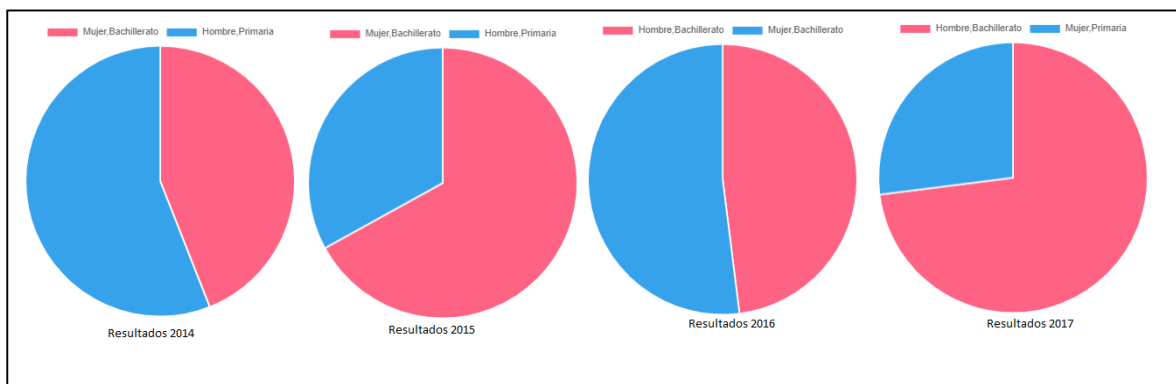


Figura 103. Resultados educación de la madre en pruebas saber 9°
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En los resultados del nivel de formación de la madre se evidencia que para el año 2014 las madres de las niñas logran terminar su bachiller, mientras que las de los niños solo logran culminar su primaria, en cuanto al 2015 se evidencia la misma clasificación por género del año 2014, para el 2016 los resultados son equitativos para ambos géneros, y por último en el 2017 el nivel máximo de educación alcanzado por las madres de las niñas es primaria mientras que la de los niños es bachiller.

6.3.3. Acceso a computador

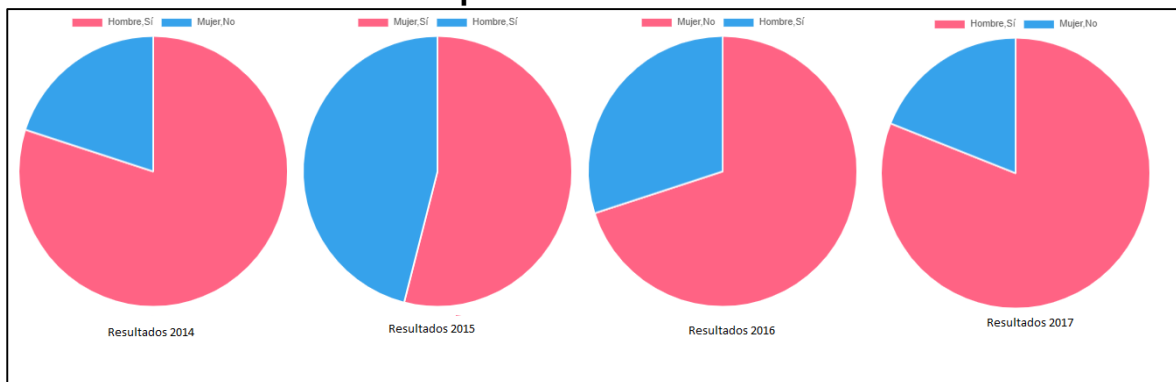


Figura 104. Resultados de acceso a computador en pruebas saber 9°
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En el acceso al computador se demuestra que para los años 2014, 2016 y 2017 las niñas dicen no tener acceso a esta herramienta, mientras que los niños si tienen acceso; lo cual corrobora la brecha de genero existente desde edades tempranas en el acceso a aparatos tecnológicos.

6.3.4. Acceso a internet

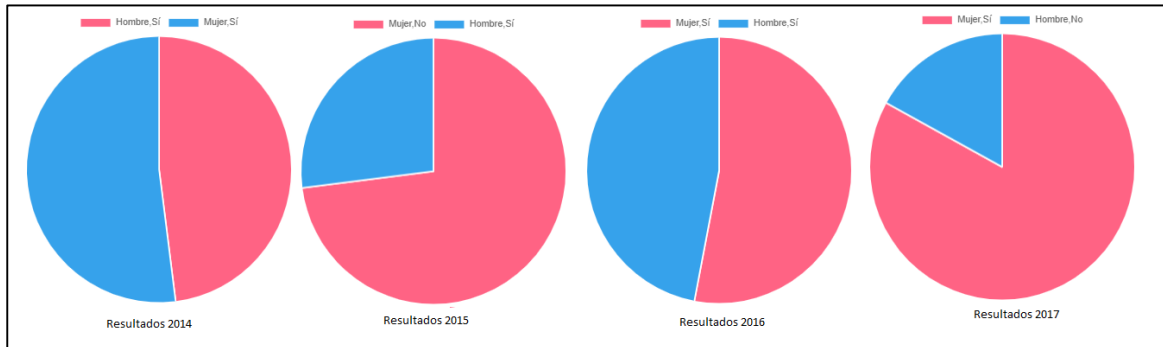


Figura 105. Resultados de acceso a internet en pruebas saber 9°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto al acceso al internet para los años 2014 y 2016 es equitativo para ambos géneros, en cambio la brecha es atribuida para el 2015 a las niñas por no tener acceso y en el 2016 a los niños por no tener; lo cual corrobora que a pesar de todo aún hay diferencias de género en el acceso al internet, provenientes de otros factores importantes.

6.3.5. Acceso a consola de video juegos

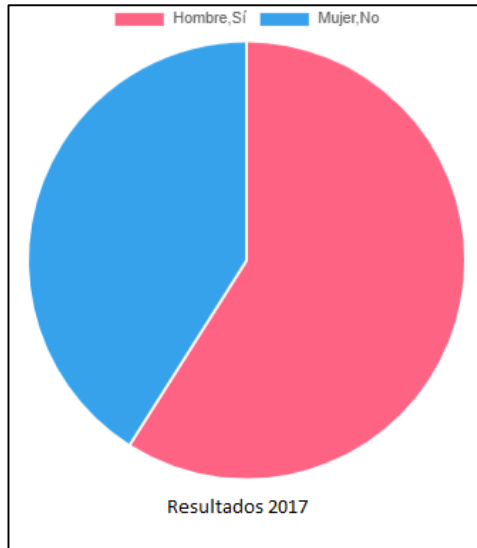


Figura 106. Resultados de acceso a consola de video-juegos en pruebas saber 9°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Con este indicador se ratifica la brecha que existe desde saber quinto en el acceso a la consola de videojuegos, ya que como se evidencia las niñas no tienen acceso; algunos autores corroboran que existen paradigmas sociales sobre el uso de dicho aparato que ha impedido el uso del mismo por parte de las niñas.

6.3.6. Resultado de puntaje en lenguaje

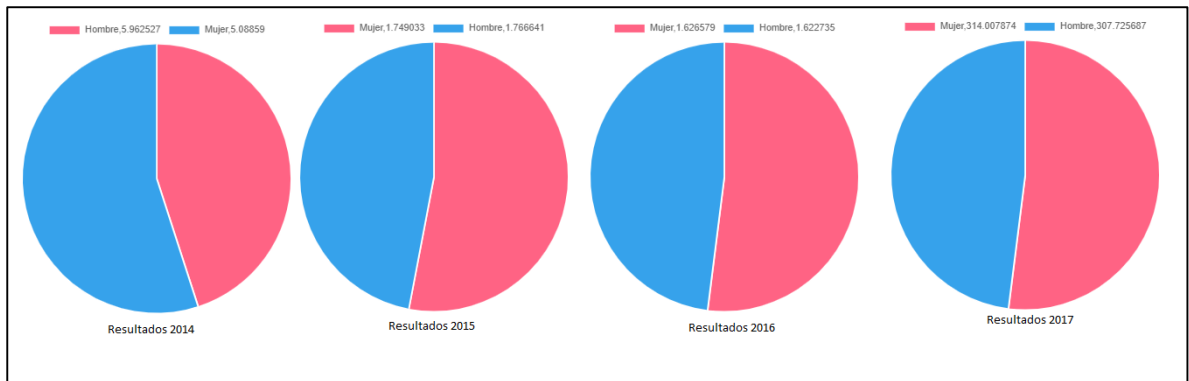


Figura 107. Resultados puntaje en lenguaje en pruebas saber 9°
Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

En cuanto al análisis de los resultados en lenguaje se evidencia que para los años 2014 y 2015 los hombres obtienen mejores resultados, mientras que en el 2016 son equitativos, ya para el 2017 las mujeres superan a los hombres en 7 puntos respectivamente. En comparación con las pruebas saber quinto los resultados son similares y en algunos años los hombres se destacan más que las mujeres.

6.3.7. Resultado de puntaje en matemáticas

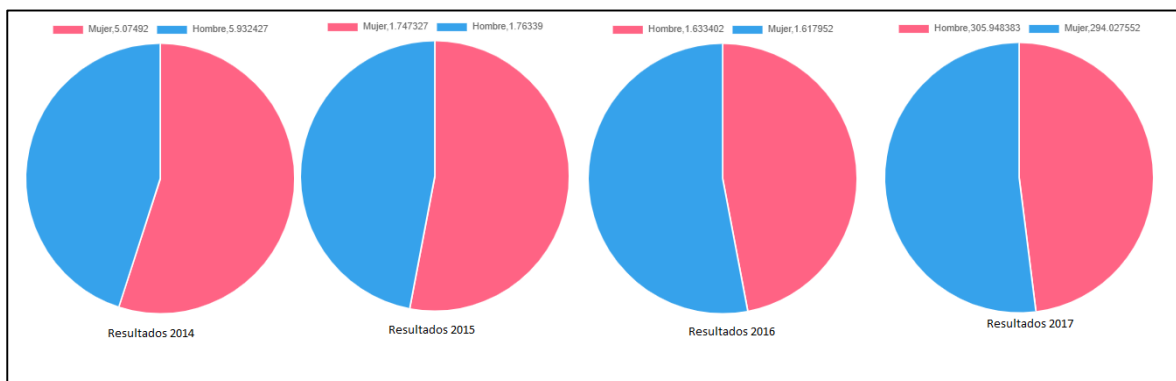


Figura 108. Resultados puntaje en matemáticas en pruebas saber 9°
 Fuente: (Viracacha Viracacha & Barahona Rodriguez, 2020)

Para los resultados en matemáticas se muestra que para todos los años, los hombres obtienen mejores puntajes, al igual que las pruebas saber quinto demostrando de esta manera la brecha existente por genero la cual se ve influenciada por aspectos tecnológicos y familiares.

CONCLUSIONES Y RECOMENDACIONES

Se evidencia que para la educación básica aún existen brechas sociales las cuales están enmarcadas por aspectos familiares tales como la educación de los padres, ya que para algunos autores este es un factor fundamental en el desarrollo de las competencias de los niños y así mismo, permite crear en el niño el modelo que desea seguir en un futuro. Observando los resultados se pudo evidenciar que la mayoría de los niveles de educación de los padres se centra en bachiller y primaria, lo cual indica que en materia de educación el país presenta falencias, ya que desde edades tempranas no se incentiva al crecimiento personal y profesional.

En el caso de acceso a tecnologías se evidencia que en la mayoría de casos las niñas aún están relacionadas con el paradigma de que los hombres son los únicos que tienen acceso a dichos aspectos, mientras ellas se dedican a las labores del hogar.

En cuanto al proceso de automatización de descarga de resultados del FTP del ICFES se puede concluir que debido a la inconsistencia y los problemas de conexión con el, no se pudo realizar dicho proceso y se optó por descargar la información manualmente, debido a que lo descargado por la herramienta presentaba fallas y esta no lograba controlar errores y los pasaban por alto.

En cuanto a las pruebas realizadas al algoritmo se puede concluir que el algoritmo J48, es eficiente y sus resultados son adecuados y concisos con lo requerido, igualmente se demostró que el error de predicción para este algoritmo fue en promedio de un 0.18, con respecto a los demás algoritmos predictivos.

Se recomienda para estudios futuros, realizar un análisis comparativo entre los resultados de las pruebas PISA y los resultados del ICFES esto para determinar

cómo es el comportamiento tanto nacional como internacional, y así mismo determinar que otros factores inciden o como los indicadores hallados inciden en las pruebas internacionales.

Igualmente se recomienda brindar a partir de los resultados, posibles estrategias que mitiguen las brechas existentes y mejoren los aspectos sociales en donde aún persisten ciertas diferencias sociales.

BIBLIOGRAFÍA

- Abadía, L. K., & Bernal, G. (2016). *Brechas de género en el rendimiento escolar a lo largo de la distribución de puntajes: evidencia pruebas Saber 11*. 16(9), 1–34. <https://cea.javeriana.edu.co/investigacion-publicaciones/documentos-trabajo/vniversitas-economica>
- Ballesteros, A., Sánchez-Guzmán, D., & García, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. (Spanish). *Latin-American Journal of Physics Education*, 7(4), 662–668. <http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=94988767&lang=es&site=ehost-live>
- Barrientos, R., Cruz, N., Acosta, H., Rabatte, I., Gogeoascoechea, M., Pavón, P., & Blázquez, S. (2009). *Arbol de decisión como herramienta en el diagnostico medico*. 20–24. https://www.uv.mx/rm/num_anteriores/revmedica_vol9_num2/articulos/arboles.pdf
- Camejo, I. M., Medina, G. A., Beltrán, P. M., & Carrillo, K. G. (2015). *Desempeño de algoritmos de minería en indicadores académicos : Árbol de Decisión y Regresión Logística Performance of data mining algorithms in academic indicators : Decision Tree and Logistic Regression*. 9(4), 104–116
- Congreso de la república ley número 23 de 1982. (1982).
- Congreso de la república ley número 565 de 2000 (2000).
- Congreso de la república ley número 1712 de 2014 (2014).
- Congreso de la república Decreto Número 1360 de 1989 (1989).
- Constitución política de Colombia. 1991. Gaceta Constitucional No. 116 de 20 de julio de 1991.
- D. Martínez, M. Karanik, M. Giovannini, N. P. (2015). Perfiles de Rendimiento Académico : Un Modelo basado en Minería de datos Academic Performance Profiles : A Model based on data Mining. *Campus Virtuales*, IV(2015), 12–30. <http://uajournals.com/ojs/index.php/campusvirtuales/article/view/66/65>
- Gomez, L. M., & Jaramillo, Y. A. (2017). DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO EN LAS PRUEBAS SABER 5 CON TÉCNICAS DESCRIPTIVAS DE MINERÍA DE DATOS. *Ekp*, 13(3), 1576–1580.
- Grijalva Arriaga, P. K., Freire Aviles, V., Real Aviles, K., & Arellano Arcentales, A. (2018). *Rev. Hallazgos21, Vol. 3, Suplemento Especial, 2018 TÉCNICAS DE MINERÍA DE DATOS EN LA EFICIENCIA ACADÉMICA*. 3, 1–16. <https://revistas.pucese.edu.ec/hallazgos21/article/view/222>
- Hernández-Sampieri, R., Fernández-Collado Carlos, & Baptista-Lucio, P. (2006). Los enfoques cuantitativo y cualitativo en la investigación científica. In *Metodología de la investigación*. <https://doi.org/10.6018/turismo.36.231041>
- López, V., Grazi, M., Guillard, C., & Salazar, M. (2018). Las brechas de género en

- ciencia, tecnología e innovación en América Latina y el Caribe. *Banco Interamericano de Desarrollo*, 52. <https://doi.org/10.18235/0001082>
- Menacho Chiok, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26. <https://doi.org/10.21704/ac.v78i1.811>
- Merchán Rubiano, S., Beltrán Gómez, A., & Duarte García, J. (2017). Engineering Students' Academic Performance Prediction using ICFES Test Scores and Demo-graphic Data. *Ingeniería Solidaria*, 13(21), 53–61. <https://doi.org/10.16925/in.v13i21.1729>
- Mineducación. (2013). Colombia en PISA 2012, principales resultados. *Ministerio de Educación Nacional Republica Colombia*, 1–44. http://www.icfes.gov.co/2012-07-05-14-55-31/doc_download/8820-presentacion-principales-resultados-colombia-en-pisa-2012%5Cnhttp://www.icfes.gov.co/2012-07-05-14-55-31/doc_download/5882-saber-pro-principales-resultados-en-competencias-genericas-2011.pdf
- Moine, J. M., Gordillo, S., & Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. *Xvii Congreso Argentino De Ciencias De La Computación*, 931–938.
- Moine, J. Mi., Haedo, A., & Gordillo, S. (2011). Estudio comparativo de metodologías para minería de datos. *XIII Workshop de Investigadores En Ciencias de La Computación*, 278–281. <http://sedici.unlp.edu.ar/handle/10915/20034>
- Moreno, J., Rodríguez, D., Sicilia, M. A., Riquelme, J. C., & Ruiz, R. (2009). *SMOTE-I: mejora del algoritmo SMOTE para balanceo de clases minoritarias*. 3(1), 73–80.
- Ojeda García, A. (2014). Los retos de la mujer del siglo XXI, en materia de educación para impulsar una perspectiva de género. *Aloma: Revista de Psicología, Ciències de l'educació i de l'esport*, 32(32), 65–75.
- Oviedo Carrascal, A. I., & Jiménez Giraldo, J. (2019). Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas SABER-PRO. *Revista Politécnica*, 15(29), 128–140. <https://doi.org/10.33571/rpolitec.v15n29a10>
- Ricardo Timarán-Pereira, R., Caicedo-Zambrano, J., & Hidalgo-Troya, A. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Revista De Investigación, Desarrollo E Innovación*, 9(2), 363–378. <https://doi.org/10.19053/20278306.v9.n2.2019.9184>
- Rico Paez, A., & Sanchez Guzman, D. (2018). Análisis de datos de estudiantes de ingeniería para la predicción del rendimiento académico mediante técnica de clasificación bayesiana. (April), . <https://doi.org/10.1037/0033-2909.126.1.78>
- Rico Páez, A., & Sánchez Guzmán, D. (2018). Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN / Design of a model to automate the prediction of academic performance in students of IPN. *RIDE Revista Iberoamericana Para La Investigación y El*

- Desarrollo Educativo*, 8(16), 246–266. <https://doi.org/10.23913/ride.v8i16.340>
- Rodríguez Suárez, Y., & Amador, A. D. (2009). Herramientas de Minería de Datos Data Mining Tools. *Revista Cubana de Ciencias Informaticas*, 3(3), 73–80. [https://rcci.uci.cu/?journal=rcci&page=article&op=viewFile&path\[\]=78&path\[\]=70](https://rcci.uci.cu/?journal=rcci&page=article&op=viewFile&path[]=78&path[]=70)
- Rodríguez, G. S. (1992). Revista de derecho (Baranquilla). *Revista de Derecho*, 30, 164–169. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-86972008000200007
- Sarría, F. A. (n.d.). *Programación en SQL con PostgreSQL*. 1–23.
- Singhal, S., & Jena, M. (2013). A Study on WEKA Tool for Data Preprocessing , Classification and Clustering. *International Journal of Innovative Technology and Exploring Engineering*, 2(6), 250–253. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.687.799&rep=rep1&type=pdf>
- Timaran-Pereira, R., Calderón-Romero, A., & Hidalgo-Troya, A. (2017). Aplicación de los árboles de decisión en la identificación de patrones de lesiones fatales por causa externa en el municipio de Pasto, Colombia. *Universidad y Salud*, 19(3), 388. <https://doi.org/10.22267/rus.171903.101>
- Timaran, R., & Jiménez, J. (2014). Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM. *Formacion Universitaria*, 1–19.
- Valle, U., Fonnegra, C., Byron, J., & Valle, U. (2016). Desempeño académico y diferencias de género en Colombia: un análisis con base en las pruebas TIMSS 2007. *Revista Sociedad y Economía*, 30, 15–42.

ANEXOS

ANEXO A Manual de usuario



UDEC
UNIVERSIDAD DE
CUNDINAMARCA

APLICATIVO WEB PARA LA CONSTRUCCIÓN DE
INDICADORES DE BRECHA DE GÉNERO EN LA
EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL PAÍS



MANUAL DE USUARIO
DESARROLLO DE SOFTWARE
UNIVERSIDAD DE CUNDINAMARCA
2020

www.unicundi.edu.co
unicundi@mail.unicundi.edu.co
Línea gratuita 018000 976000



GP-CER355041

CO-SC-CER355037

SC-CER355037

Dirección de Sistemas y Tecnología
sistemasytecnologia@mail.unicundi.edu.co
PBX: 828 14 83 Ext. 110-170
Sede Fusagasugá

Institución de educación superior sujeta a inspección y vigilancia por el Ministerio de Educación Nacional



TABLA DE CONTENIDO

INTRODUCCIÓN	179
1. USUARIOS	180
1.1. ADMINISTRADOR	180
1.2. USUARIO	180
2. REQUISITOS DE SOFTWARE Y HARDWARE	181
2.1. REQUERIMIENTOS DE SOFTWARE	181
3. APLICATIVO	182
3.1. ICONOS GENERALES	182
3.2. INGRESO AL APLICATIVO	182
3.2.1. Inicio de sesión	183
3.2.2. Contáctenos	183
3.2.3. MÓDULO ADMINISTRADOR	184
3.2.3.1. Página principal del módulo administrativo	184
3.2.3.2. Insertar usuarios	188
3.2.3.3. Listado de usuarios	189
3.2.3.4. Listado de pruebas	191
3.2.3.5. Cargar archivos pruebas	191
3.2.4. MÓDULO DE USUARIOS	195
3.2.4.1. Página principal del módulo de usuarios	195
3.2.4.2. Resultados descriptivos saber tercero	197
3.2.4.3. Resultados descriptivos saber quinto y noveno	199
3.2.4.4. Resultados predictivos saber básico	201
CONTROL DE CAMBIOS DEL MANUAL	205



TABLA DE IMÁGENES

Imagen 1. Página principal del aplicativo	182
Imagen 2. Página de inicio de sesión	183
Imagen 3. Formulario contáctenos	184
Imagen 4. Mensaje sobre el estado de la solicitud	184
Imagen 5. Página principal del módulo administrativo	185
Imagen 6. Lista de correos recibidos	185
Imagen 7. Lista de notificaciones de archivos	186
Imagen 8. Botón página principal	186
Imagen 9. Menú barra lateral izquierda	187
Imagen 10. Contenido de la página principal	188
Imagen 11. Registro de usuarios	188
Imagen 12. Correo con datos de acceso	189
Imagen 13. Listado de usuarios	189
Imagen 14. Editar usuario	190
Imagen 15. Mensaje de edición exitosa	190
Imagen 16. Mensaje de eliminación exitosa	190
Imagen 17. Listado de pruebas cargadas al aplicativo	191
Imagen 18. Formulario inicial de carga de pruebas	191
Imagen 19. Selección tipo de prueba y año	192
Imagen 20. Carga de pruebas del año 2016 hacia atrás.	192
Imagen 21. Carga de pruebas del año 2016 hacia adelante.	193
Imagen 22. Panel de ayuda.	193
Imagen 23. Vista de página notificaciones de carga de archivos	194
Imagen 24. Vista corrección de carga de archivos	194
Imagen 25. Barra superior del módulo de usuarios	195
Imagen 26. Formulario de envío de mensajes	195
Imagen 27. Menú lateral izquierdo módulo de usuarios	196
Imagen 28. Contenido de página principal de modulo usuarios	197
Imagen 29. Vista análisis de pruebas saber 3	197
Imagen 30. Graficas análisis saber tercero	198
Imagen 31. Pagina de análisis saber quinto	199
Imagen 32. Pagina de análisis saber noveno	199
Imagen 33. Primera parte de graficas respecto al genero	200
Imagen 34. Segunda parte de graficas respecto al indicador	200
Imagen 35. Tercera parte de graficas: relación de indicadores	201
Imagen 36. Vista inicial resultados predictivos saber básico	201
Imagen 37. Vista grafica de los resultados predictivos	202
Imagen 38. Filtros del costado izquierdo	202
Imagen 39. Filtro de indicador	203
Imagen 40. Resultados predictivos por género	203
Imagen 41. Detallado del indicador	203



INTRODUCCIÓN

En este documento se explicara el funcionamiento y el manejo del aplicativo web para la construcción de indicadores de brecha de género en la educación básica del país, aquí se evidenciaran las funcionalidades de acuerdo a los roles existentes.

El proyecto como tal busca identificar qué factores influyen en la brecha de género del país, teniendo como referencia los resultados de las pruebas saber 3, 5 y 9 respectivamente; por ello el resultado de dichos análisis es un sitio web en el cual el usuario podrá visualizar los resultados hallados de manera gráfica y en reportes más detallad



1. USUARIOS

1.1. ADMINISTRADOR

El administrador es el usuario encargado del manejo administrativo del aplicativo web, sus funciones son:

- Registrar, modificar, consultar y eliminar usuarios
- Cargar pruebas saber y llevar un control sobre las pruebas que presentan fallos.
- Ver listado de pruebas cargadas en base de datos
- Ver notificaciones de correos y fallo en pruebas cargadas

1.2. USUARIO

El usuario es el rol que tiene acceso a los resultados de los análisis hechos con las pruebas sus funciones son:

- Ver resultados descriptivos de las pruebas saber 3, 5 y 9
- Ver resultados predictivos de las pruebas saber 3, 5 y 9



2. REQUISITOS DE SOFTWARE Y HARDWARE

2.1. REQUERIMIENTOS DE SOFTWARE

- Contar con un equipo electrónico, ya sea computador o portátil.
- Navegador: Contar con navegadores como Firefox o Google Chrome.
- Es indispensable contar con acceso a internet.





3. APLICATIVO

APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE BRECHA DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL PAÍS

3.1. ICONOS GENERALES

Tabla 1. Iconos generales

	Logo principal del aplicativo web.
	Botón de ayuda.

3.2. INGRESO AL APLICATIVO

Para ingresar al aplicativo web es necesario ingresar a la siguiente dirección ingapge.ddns.net/Ingapge, allí se visualizara la siguiente ventana:

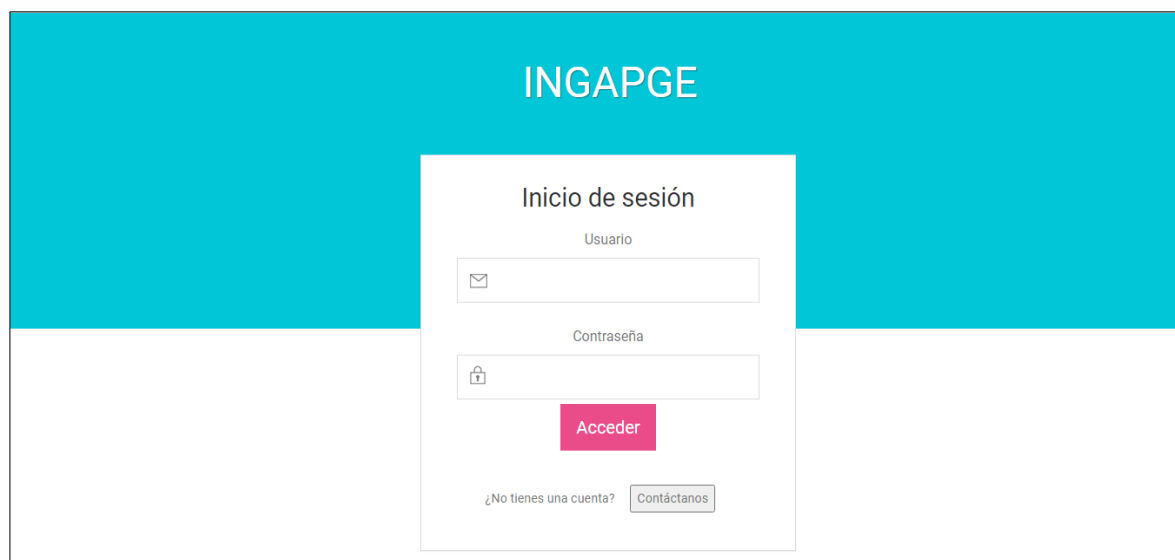


Imagen 1. Página principal del aplicativo

La imagen anterior muestra la página principal del aplicativo web, la cual corresponde al inicio de sesión.



3.2.1. Inicio de sesión

Para iniciar la persona debe acceder con el usuario y la contraseña asignados durante el registro.

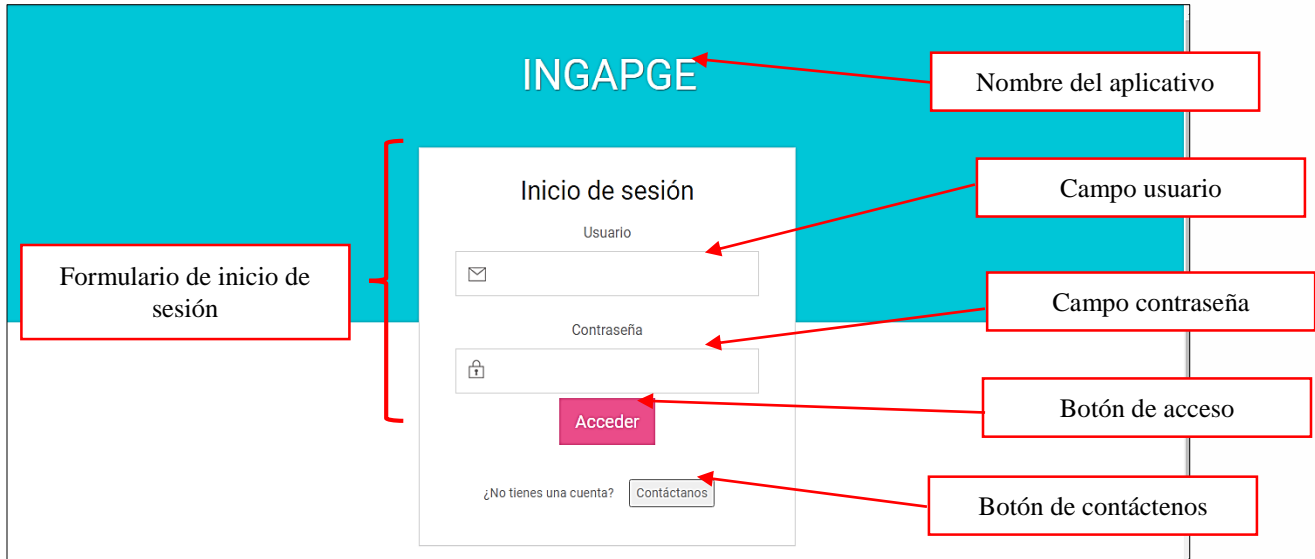


Imagen 2. Página de inicio de sesión

Si la persona no está registrada dentro del aplicativo, debe dar clic en el botón contáctenos.

3.2.2. Contáctenos

En caso de que la persona desee acceder al modulo de usuarios, debe diligenciar este formulario. Al dar clic en enviar, el administrador del aplicativo recibira un correo con los datos de contacto. Este formulario tambien es usado en caso de inconvenientes con el aplicativo, por ejemplo olvido de contraseña.



INGAPGE

Contáctenos

Nombre

Apellido

email@ucundinamarca.edu.co

Mensaje

Enviar

Formulario

Botón enviar

Título del formulario

Campo nombre

Campo apellido

Campo correo electrónico

Campo mensaje

Botón regresar al inicio de sesión

Imagen 3. Formulario contáctenos

El botón que se encuentra en la parte inferior-derecha, permite volver a la página principal.

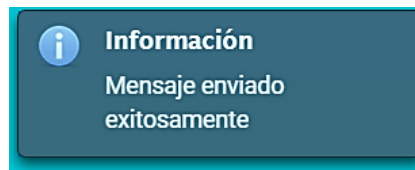


Imagen 4. Mensaje sobre el estado de la solicitud

Este mensaje indica cuando el correo se ha enviado satisfactoriamente.

3.2.3. MÓDULO ADMINISTRADOR

En este módulo gestiona la administración de los usuarios, el proceso de carga pruebas y la consulta de las pruebas cargadas al aplicativo.

3.2.3.1 Página principal del módulo administrativo

En esta página se visualizara la cantidad de registros existentes por prueba.

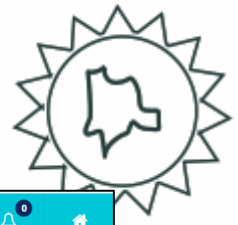


Imagen 5. Página principal del módulo administrativo

En la barra superior se encuentran tres iconos, los cuales hacen referencia a:

- El primer icono hace referencia al listado de correos recibidos, encima del icono aparece la cantidad de correos recibos.

Botón listado de correos recibidos



Imagen 6. Lista de correos recibidos

- El segundo icono hace referencia a las notificaciones de los estados en cuanto a la carga de pruebas saber 3, 5 y 9

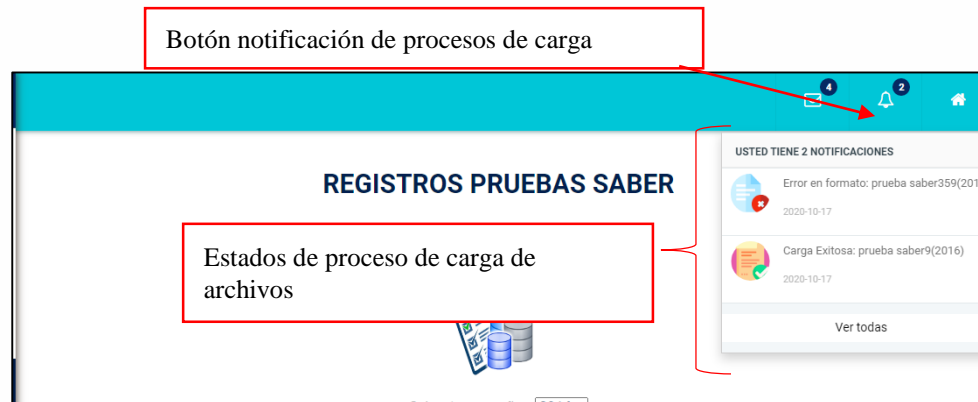


Imagen 7. Lista de notificaciones de archivos

Para este aspecto se tienen en cuenta cuatro estados:

1. Carga exitosa, cuando el archivo ha sido cargado a la base de datos exitosamente
 2. Error en formato, cuando el formato no cumple con los requisitos
 3. Error en contenido, cuando en el archivo existen más variables que columnas
 4. Error de archivo, cuando el archivo presenta fallas a la hora de leer.
 - 5.
- El tercer icono se usa para volver a la página principal del modulo

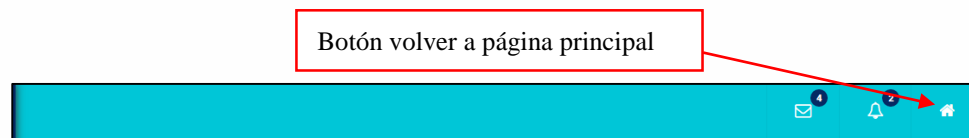


Imagen 8. Botón página principal



En el menú de la parte izquierda encontramos:



Imagen 9. Menú barra lateral izquierda

- Al dar clic en el botón para salir del módulo, este redireccionará al inicio de sesión.
- El botón de reducir menú permite expandir o contraer el menú

En cuanto al contenido de la página principal, se encuentra la cantidad de registros por cada prueba saber cargada, en dicha vista se pueden realizar filtros por año.



Prueba	Número de registros
Saber 3	355670
Saber 5	350491
Saber 9	295410

Imagen 10. Contenido de la página principal

3.2.3.2 Insertar usuarios

En este formulario se debe ingresar los datos del usuario tales como nombre, apellido, correo y contraseña, todos estos campos serán validados dentro del aplicativo.

Nombre

Apellido

Correo

Contraseña

Imagen 11. Registro de usuarios

Una vez que se ingresen los datos y se dé clic sobre el botón agregar, al usuario le llegara un correo el usuario y la contraseña de acceso.

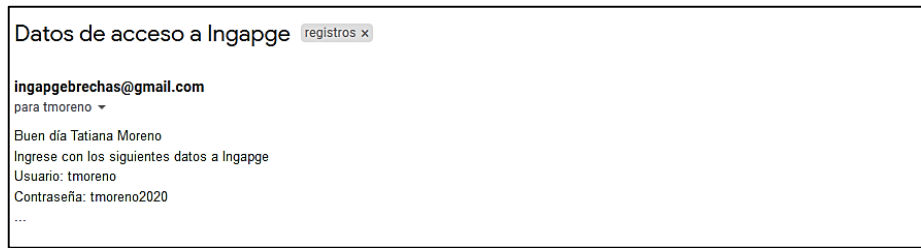


Imagen 12. Correo con datos de acceso

3.2.3.3 Listado de usuarios

En esta página se observa el listado de usuarios registrados en el aplicativo, asimismo permite editar o eliminar información de los usuarios.

La tabla muestra los siguientes datos del usuario: nombre, apellido, correo institucional, usuario y contraseña.

Nombre	Apellido	Correo Institucional	Usuario	Contraseña	
Corin	Vira	cviracacha@ucundinamarca.edu.co	cviracacha	Corin9806	
Tatiana	Gomez	tmoreno@ucundinamarca.edu.co	tmoreno	2020*10	
Juana	Gomez	gomez@gmail.com	Isa122	12sasd5	
Angie	Manrique	apaolamanrique@ucundinamarca.edu.co	apaolamanrique	Apaola1234	
Corin	Vira	cviracacha@ucundinamarca.edu.co	cviracacha	Corin9806	
Sara	Useche	nijacovi@gmail.com	nijacovi	sarita2015	
Tatiana	Moreno	tmoreno@ucundinamarca.edu.co	tmoreno	tmoreno2020	
Admin	Admin	apaolamanrique@ucundinamarca.edu.co	adminingapge	4DM1N1STR4R_2020	

Imagen 13. Listado de usuarios

En el caso de seleccionar el botón de editar, se habilitaran las casillas para modificar los datos del usuario, solo se puede editar el nombre, apellido y contraseña.



Botón editar



Activación de campos

Nombre	Apellido	Correo Institucional	Usuario	Contraseña	
Corin	Vira	cviracacha@ucundinaman	cviracacha	Corin9806	<input type="checkbox"/> <input type="checkbox"/>
Taliana	Gomez	tmoreno@ucundinamarca.edu.co	tmoreno	2020*10	<input type="checkbox"/> <input type="checkbox"/>
Juana	Gomez	gomez@gmail.com	Isa122	12sasd5	<input type="checkbox"/> <input type="checkbox"/>

Confirmar o cancelar modificación

Imagen 14. Editar usuario



Botón confirmar modificación



Botón cancelar modificación

En caso de que el proceso de modificación sea exitoso el aplicativo mostrara el siguiente mensaje:

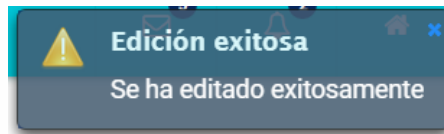


Imagen 15. Mensaje de edición exitosa

En el caso de seleccionar el botón eliminar, el usuario será eliminado del aplicativo



Botón eliminar usuario

En caso de que el proceso de eliminación sea exitoso el aplicativo mostrara el siguiente mensaje:

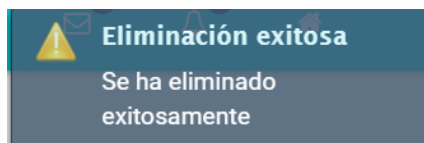


Imagen 16. Mensaje de eliminación exitosa



3.2.3.4 Listado de pruebas

En esta página se puede visualizar las pruebas que han sido cargadas al aplicativo web.

Año	Saber 3	Saber 5	Saber 9	Saber 11	Saber TyT	Saber Pro
2014	✓	✓	✓			
2015	✓	✓	✓			
2016	✓	✓	✓	✓	✓	✓
2017	✓	✓	✓	✓	✓	✓
2018				✓	✓	✓
2019				✓	✓	✓

Imagen 17. Listado de pruebas cargadas al aplicativo

3.2.3.5 Cargar archivos pruebas

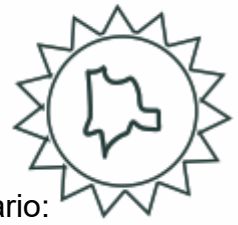
En esta vista se debe el tipo de prueba a cargar, ya sean resultados pruebas saber 11, TyT y Pro o resultados 3,5 y 9.

Datos pruebas

Resultados 3, 5 y 9 Resultados 11, TyT y Pro

Selector tipo de educación

Imagen 18. Formulario inicial de carga de pruebas



Una vez seleccionado “Resultados 3, 5 y 9” aparecerá el siguiente formulario:

Carga resultados Saber 3, 5 y 9

Prueba a cargar:

Año de presentación de la prueba:

Continuar

Selección prueba

Selección año

Selector tipo de prueba

Botón de ayuda

Botón de procesos pendientes

Botón continuar

Selector del año

Imagen 19. Selección tipo de prueba y año

En cuanto a los tipos de prueba, se cuenta con:

- Saber 3
- Saber 5
- Saber 9
- Todas, esta se debe porque desde el año 2017 en adelante el ICFES carga un solo archivo con los resultados de las tres pruebas.

Estos datos dependen de las pruebas y años que fueron cargados anteriormente. Al dar clic en el botón continuar, el aplicativo solicitará la carga de pruebas comprimidas en formato “.Zip”.

Si el año de selección es menor o igual al 2016, el aplicativo solicitará dos archivos el primero con los puntajes y el segundo con la información socioeconómica.

CARGAR ARCHIVOS

Resultados de pruebas *(.zip)

Resultados datos socioeconómicos *(.zip)

Selección archivo Ningún archivo seleccionado

Selección archivo Ningún archivo seleccionado

Cargar archivos

Cancelar Proceso

Carga de archivo de puntajes

Carga de archivo de datos Socioeconómicos

Imagen 20. Carga de pruebas del año 2016 hacia atrás.

Botón cargar archivos

Botón cancelar proceso de carga



En caso de ser mayor al 2016, solicitara solo un archivo el cual contiene la información socioeconómica y los puntajes de las tres pruebas.

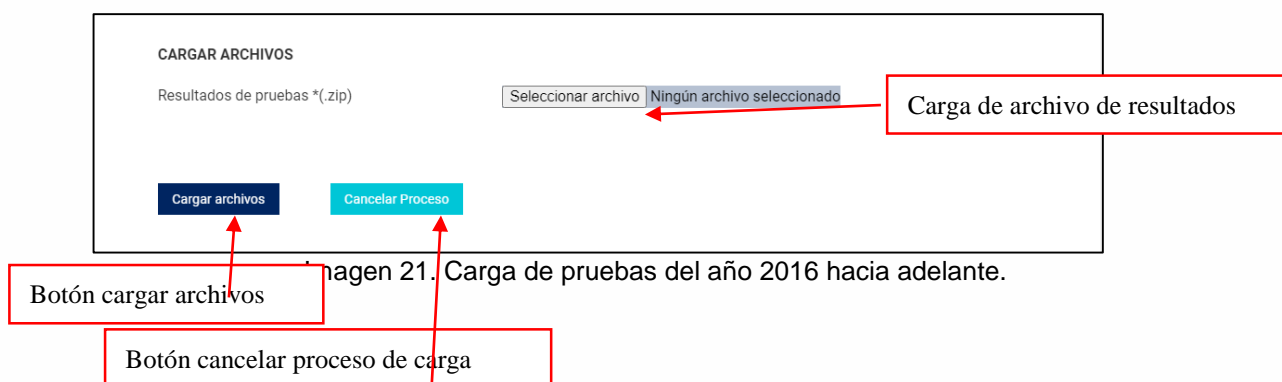


Imagen 21. Carga de pruebas del año 2016 hacia adelante.

En el caso de seleccionar el botón de cargar archivos el aplicativo se encargara de guardar los archivos en el servidor, si por el contrario selecciona el botón cancelar proceso, el aplicativo volverá a mostrar la selección de año y tipo de prueba.

- Ayudas: Al dar clic sobre el siguiente icono el aplicativo mostrara información que permita comprender el funcionamiento de la página de carga de pruebas.



Botón de ayudas de la página

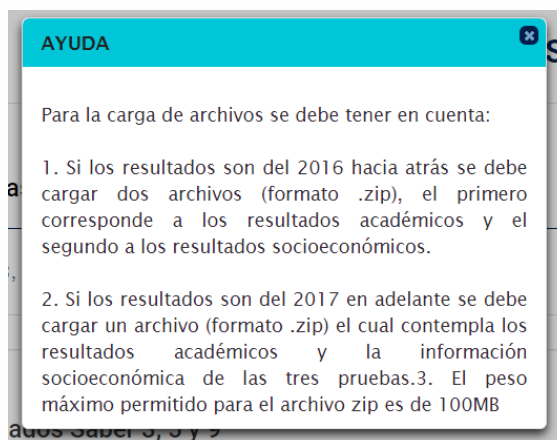


Imagen 22. Panel de ayuda.

- Notificaciones de procesos de carga de archivos: Al dar clic sobre el siguiente icono, este redireccionará a la vista de notificaciones de procesos de carga de archivos pendiente.



Botón notificaciones de procesos de carga



Vista de notificaciones, en la tabla se puede evidenciar el nombre de la prueba, año, tipo de resultados y el error del archivo.

Fecha: 19-10-2020

NOTIFICACIONES PENDIENTES

Procesos pendientes de carga de resultados

No	Prueba	Año	Tipo de resultados	Error
1	saber 359	2018	puntajes-economico	Error en formato

Botones de corregir archivo y eliminar proceso

Botón de ayudas

Listado de notificaciones

Imagen 23. Vista de página notificaciones de carga de archivos

Si selecciona el botón “eliminar proceso” se eliminara la notificación, los datos de la prueba y el archivo cargado.

Al seleccionar el botón corregir, el aplicativo mostrara el siguiente panel, el cual permite cargar nuevamente el archivo que presento errores, al dar clic sobre el botón guardar resultados el aplicativo remplazara el archivo antiguo por el corregido y vuelve a iniciar el proceso de carga a la base de datos:

Correcciones carga de resultados

Cargar resultados corregidos

Seleccionar archivo Ningún archivo seleccionado

Input, para cargar el archivo

Botón de guardar resultados o cancelar proceso

Imagen 24. Vista corrección de carga de archivos



3.2.4. MÓDULO DE USUARIOS

En este módulo se tendrá acceso a los resultados de los indicadores de brecha de género, asimismo se podrá visualizar los resultados predictivos hallados con las pruebas ICFES desde educación primaria hasta superior.

3.2.4.1. Página principal del módulo de usuarios

En esta página se puede visualizar la cantidad de mujeres inscritas por pruebas, asimismo podrá realizar filtros sobre los datos hallados.

En la barra superior se encuentran dos iconos, el primero permite volver a la página principal del módulo y el segundo redireccionará a la página de contacto



Imagen 25. Barra superior del módulo de usuarios

Una vez seleccionado el icono de mensaje aparece el envío de mensajes.

Botones para volver a la página principal y mostrar un chat

ENVIAR MENSAJES

Campo para enviar mensaje

Enviar Cancelar

Campo para enviar mensaje

Botón de enviar y cancelar proceso

Imagen 26. Formulario de envío de mensajes



En el menú de la parte izquierda encontramos:



Imagen 27. Menú lateral izquierdo módulo de usuarios

En el contenido de la página principal de este módulo, se puede visualizar la cantidad de mujeres registradas por prueba, asimismo se permite realizar filtros por año y tipo de prueba.

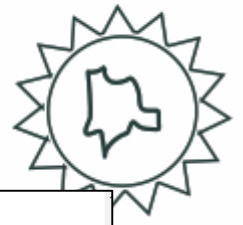


Imagen 28. Contenido de página principal de modulo usuarios

3.2.4.2. Resultados descriptivos saber tercero

Para esta página inicialmente se debe seleccionar el año y el indicador a analizar.

ANÁLISIS DE RESULTADOS SABER 3

Seleccionar año a analizar:

2015

Seleccionar indicador a analizar:

Analizar

Seleccionar indicador

Puntaje Lenguaje

Puntaje Matemáticas

Botón de ayuda

Filtro de año

Filtro del indicador

Botón analizar

Imagen 29. Vista análisis de pruebas saber 3



Al dar clic sobre el botón analizar aparecerán las siguientes gráficas, las cuales corresponden al análisis hecho

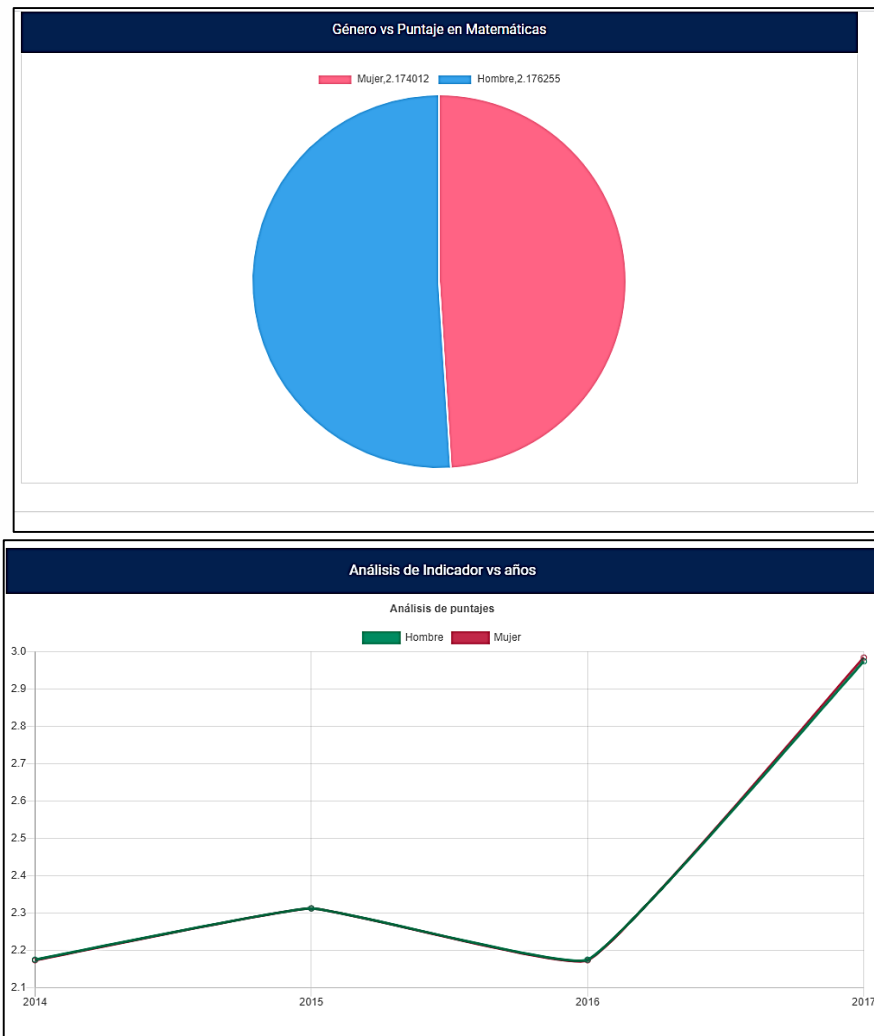


Imagen 30. Gráficas análisis saber tercero

Al final de las gráficas encontrara un botón el cual permite descargar el reporte en formato PDF.



Botón para descargar reportes en PDF.



3.2.4.3. Resultados descriptivos saber quinto o noveno

En cuanto al contenido y la cantidad de gráficas de las pruebas saber quinto son iguales a los análisis de las pruebas saber noveno.

- Primero se debe seleccionar del menú principal la prueba saber, a la cual se le quiere realizar el análisis.
- Posteriormente seleccionar el año y el indicador.

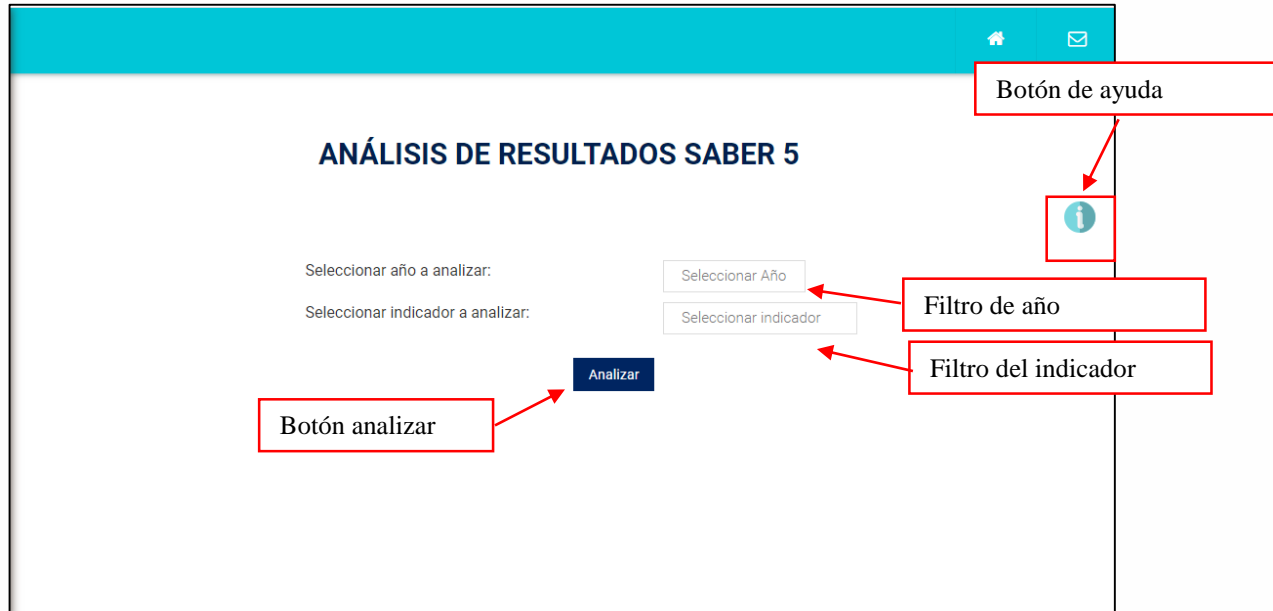


Imagen 31. Página de análisis saber quinto

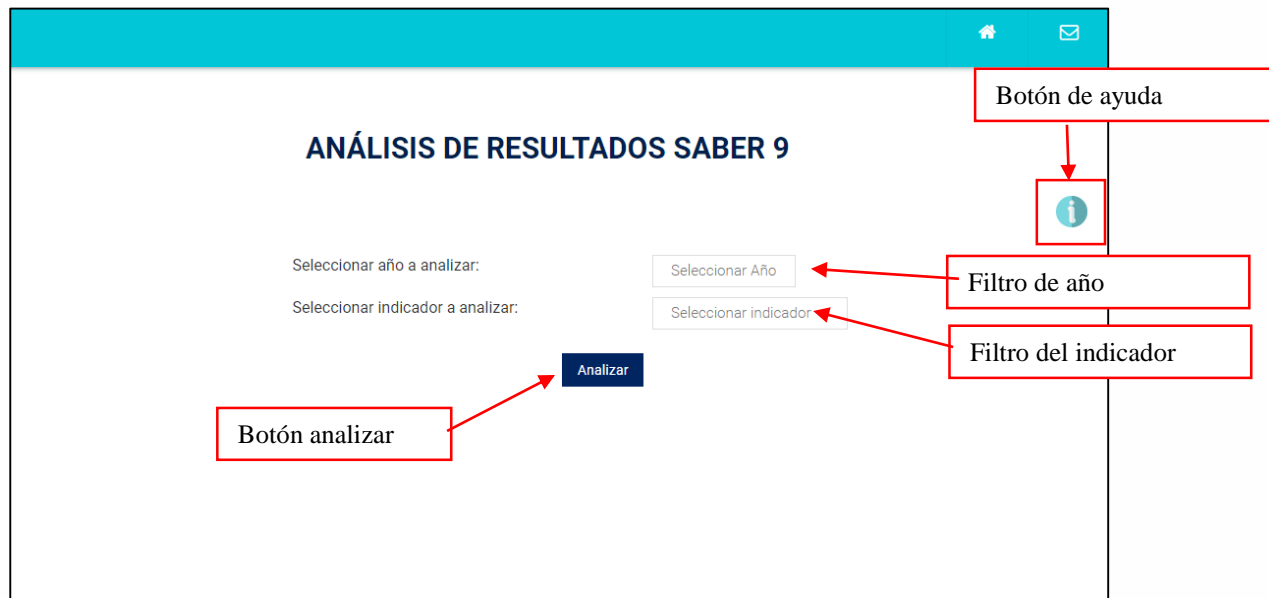


Imagen 32. Página de análisis saber noveno



Al dar clic en el botón analizar se podrán evidenciar las siguientes graficas:

- En estas dos primeras graficas se establece la relación del genero vs el indicador

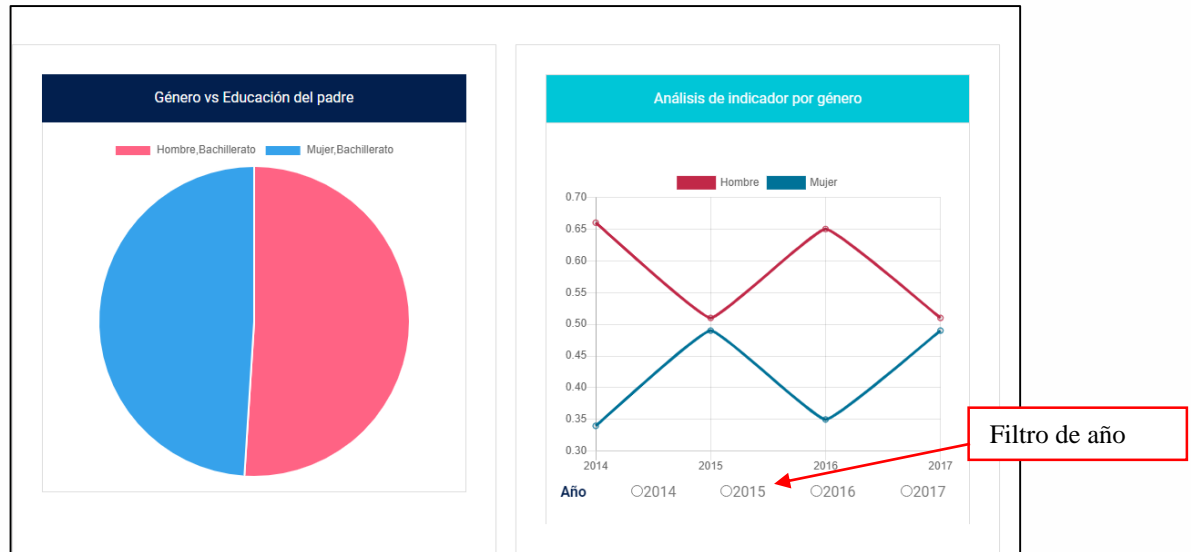


Imagen 33. Primera parte de graficas respecto al género

- En las siguientes graficas se visualiza la relación del indicador en cuanto al año y al tipo de prueba.

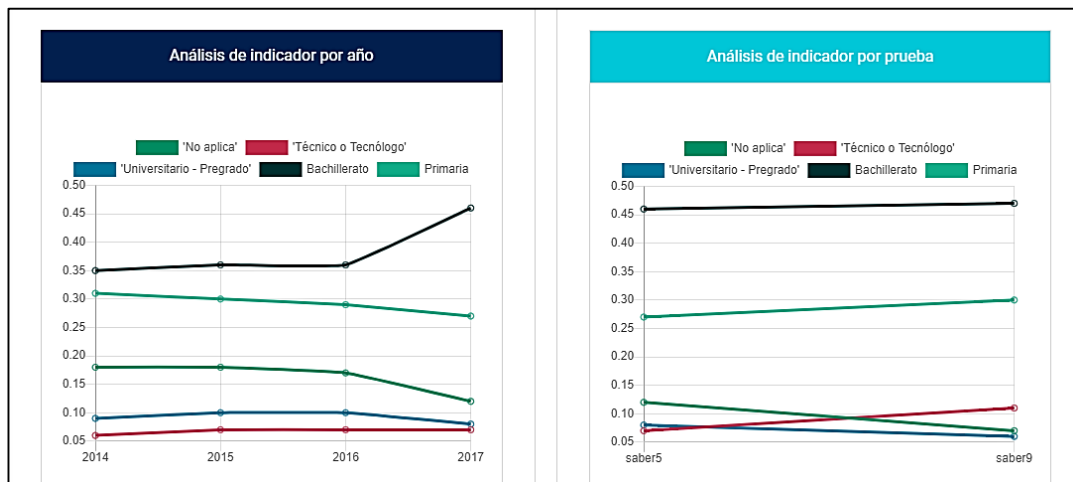


Imagen 34. Segunda parte de graficas respecto al indicador



- La última grafica hace referencia a la relación entre todos los indicadores por año. Aquí se pueden realizar consultas de los indicadores por año.

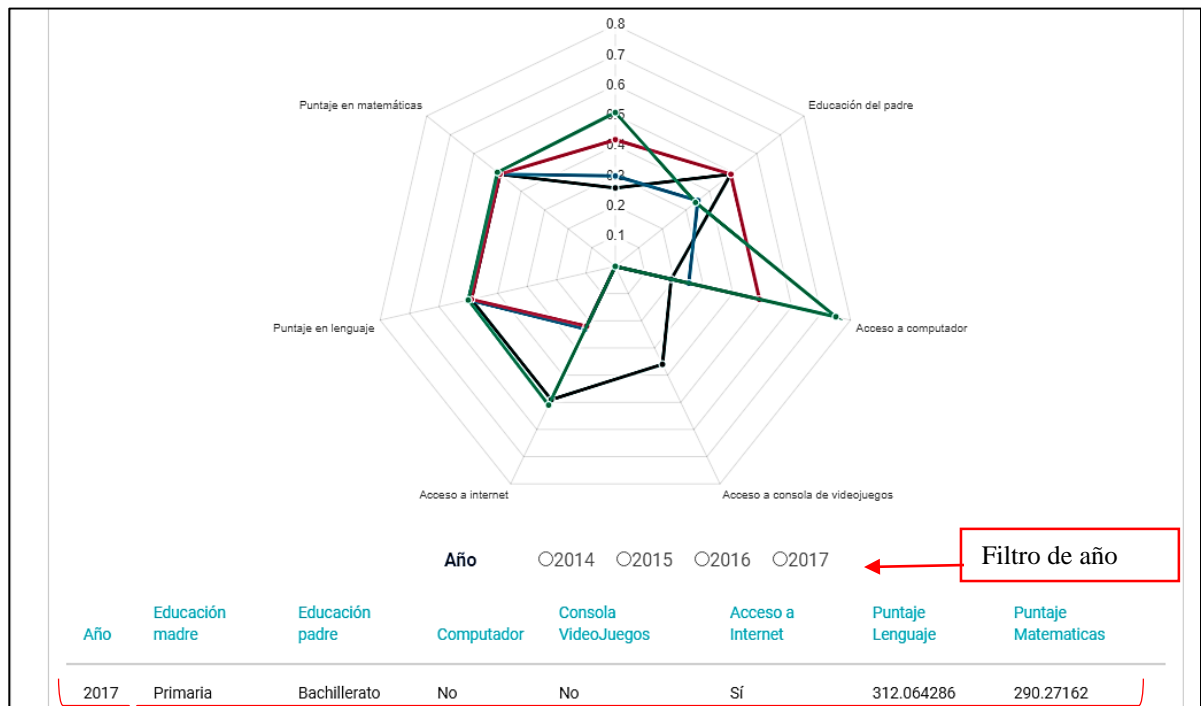


Imagen 35. Tercera parte de graficas: relación de indicadores

Información del año seleccionado

Al igual que las pruebas saber tercero al final de las gráficas hay un botón con el icono del PDF, el cual permite generar reportes de los análisis hechos.

3.2.4.4. Resultados predictivos saber básico

En la siguiente vista se podrá visualizar los resultados predictivos de las pruebas saber básico.

ANÁLISIS PREDICTIVO DE EDUCACIÓN BÁSICA

Botón de ayuda

Seleccionar por

- Saber 5
- Puntaje Matemáticas
- Saber 9
- Puntaje Lenguaje

Filtro de tipo de prueba y núcleo temático

Imagen 36. Vista inicial resultados predictivos saber básico



Una vez seleccionado el tipo de prueba y el núcleo temático el aplicativo mostrará el siguiente contenido:

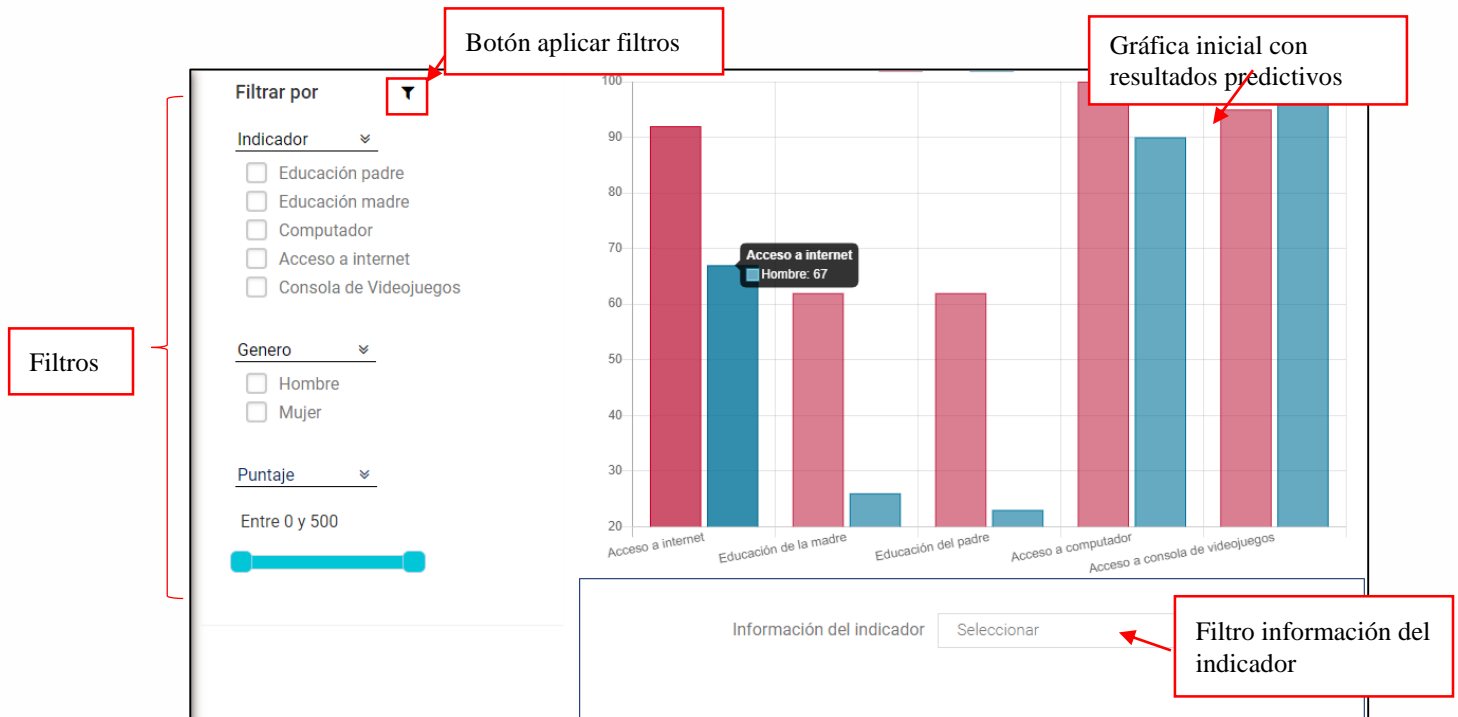


Imagen 37. Vista grafica de los resultados predictivos

Para aplicar los filtros del costado izquierdo, primero se debe seleccionar las variables a analizar y posteriormente seleccionar el botón que tiene el icono del filtro.

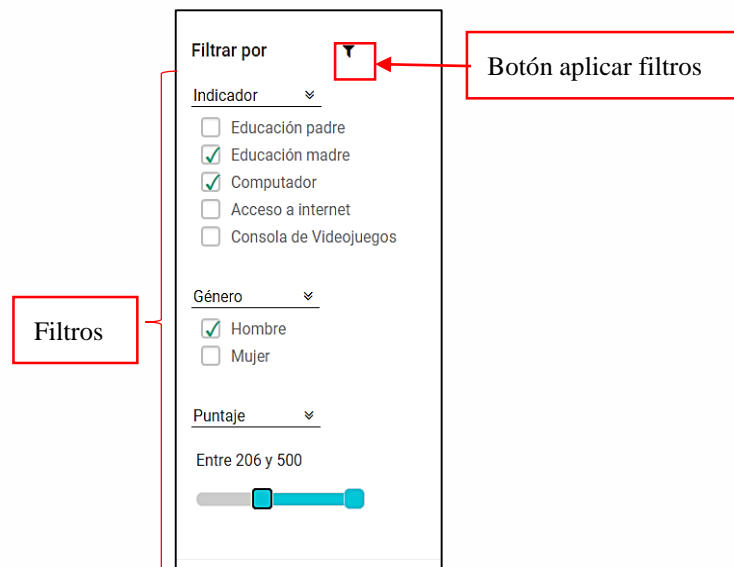


Imagen 38. Filtros del costado izquierdo



En cuanto a la información del indicador, primero se debe seleccionar el indicador a consultar, para que el aplicativo muestre los resultados predictivos con mayor influencia.

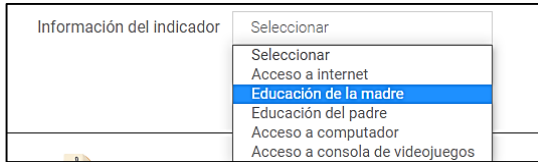


Imagen 39. Filtro de indicador

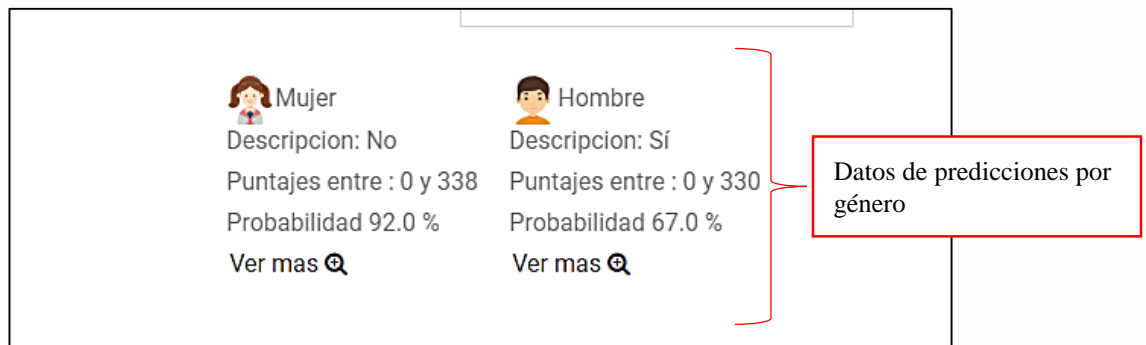


Imagen 40. Resultados predictivos por género

Al seleccionar el botón “ver más”, se puede obtener más información del indicador seleccionado.

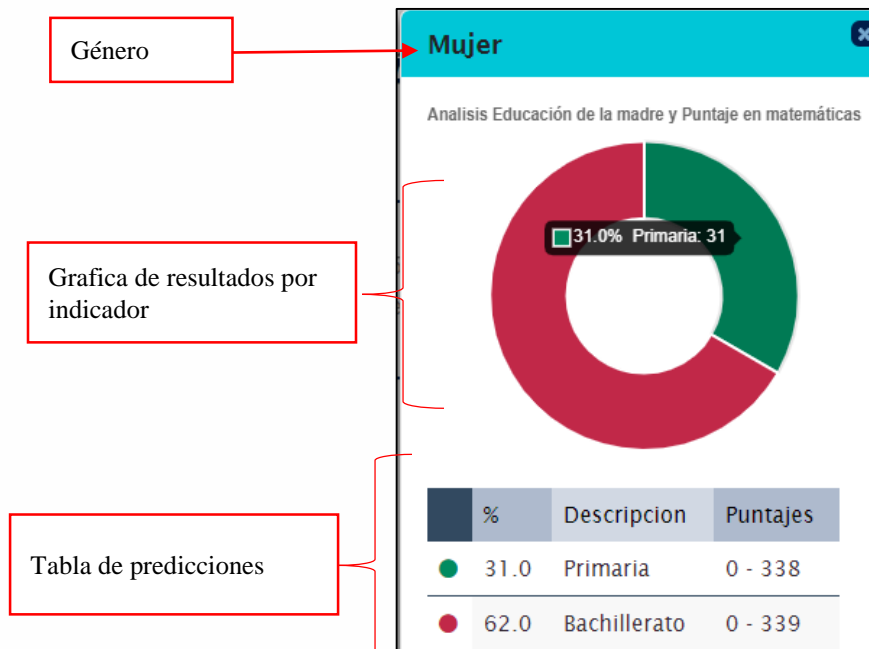


Imagen 41. Detallado del indicador



En la tabla de predicciones se puede encontrar inicialmente el porcentaje de predicción, la descripción del indicador y los posibles puntajes a obtener en dicho núcleo temático.

Al igual que los análisis descriptivos al final del formulario se encontrara un botón con el icono del PDF, el cual permite descargar un reporte más detallado sobre las predicciones echas.



Botón para generar reportes



CONTROL DE CAMBIOS DEL MANUAL

Actualización Nro.	Descripción del cambio	Versión del aplicativo	Fecha de cambio



www.unicundi.edu.co
unicundi@mail.unicundi.edu.co
Línea gratuita 018000 976000



GP-CER365641



CO-SC-CER365037



SC-CER365037

Dirección de Sistemas y Tecnología
sistemasytecnologia@mail.unicundi.edu.co
PBX: 828 14 83 Ext. 110
Sede Fusagasugá

ANEXO B Manual Técnico

**APLICATIVO WEB PARA LA CONSTRUCCIÓN DE INDICADORES DE
BRECHA DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA Y MEDIA DEL
PAÍS**

CORIN JAZMIN VIRACACHÁ VIRACACHÁ

**UNIVERSIDAD DE CUNDINAMARCA
PROGRAMA DE INGENIERÍA DE SISTEMAS
FACATATIVA
2020**

TABLA DE CONTENIDO

INTRODUCCION	210
1. ESPECIFICACIONES TÉCNICAS	211
2. INSTALACION DE SOFTWARE	212
2.1. Instalación de JDK y JRE	212
2.2. Instalación de Netbeans 8.1	217
2.3. Instalación de PostgreSQL	224
2.3.1. Instalación de extensiones Npgsql y PgAgent	228
2.4. Instalación de WEKA	230

TABLA DE IMÁGENES

Imagen 1. Página de descarga del JDK	212
Imagen 2. Ejecución de archivo JDK	212
Imagen 3. Ventana de componentes a instalar	213
Imagen 4. Proceso de extracción de archivos	213
Imagen 5. Configuración de la carpeta de destino	214
Imagen 6. Instalación de componentes	214
Imagen 7. Ventana propiedades del sistema	215
Imagen 8. Ventana variables de entorno	215
Imagen 9. Configuración inicial de la variable de entorno	216
Imagen 10. Configuración variable PATH	216
Imagen 11. Edición variable PATH	216
Imagen 12. Descarga de Netbeans	217
Imagen 13. Configuración de componentes	217
Imagen 14. Selección del servidor de despliegue	218
Imagen 15. Aceptación de términos y condiciones	218
Imagen 16. Configuración y localización de carpetas	219
Imagen 17. Información detallada de la instalación del Netbeans.	219
Imagen 18. Instalación de Netbeans	220
Imagen 19. Instalación completada de Netbeans	220
Imagen 20. Entorno de Netbeans	221
Imagen 21. Configuración de servidor de despliegue	221
Imagen 22. Añadir servidor	222
Imagen 23. Configuración de ruta de GlassFish	222
Imagen 24. Configuración del host y el puerto	223
Imagen 25. Instalación de GlassFish finalizada.	223
Imagen 26. Descarga del PostgreSQL	224
Imagen 27. Ventana principal de instalación del PostgreSQL	224
Imagen 28. Configuración de directorio de PostgreSQL	225
Imagen 29. Configuración de componentes a instalar	225
Imagen 30. Configuración de contraseña	226
Imagen 31. Configuración del puerto de conexión	226
Imagen 32. Configuración regional	227
Imagen 33. Instalación de postgresQL	227
Imagen 34. Selección del gestor de base de datos	228
Imagen 35. Selección de las extensiones a instalar	228
Imagen 36. Configuración del directorio de descarga	229
Imagen 37. Configuración del directorio de instalación de extensiones	229
Imagen 38. Descarga de WEKA	230
Imagen 39. Términos y condiciones de licencia	230
Imagen 40. Configuración en la instalación de componentes	231
Imagen 41. Configuración del directorio	231
Imagen 42. Configuración del folder inicial	232

INTRODUCCION

Este manual es una guía de instalación de los componentes y extensiones necesarias para el buen funcionamiento del aplicativo, aquí se describen los pasos necesarios que se deben cumplir para una correcta instalación y configuración. Adicionalmente con este manual se busca que el técnico comprenda más fácilmente los aspectos y características que forman parte del desarrollo del aplicativo.

1. ESPECIFICACIONES TÉCNICAS

Necesidades técnicas

- ✓ Sistema Operativo Windows 7 o superior.
- ✓ Equipo de cómputo con memoria RAM de 2GB o superior.
- ✓ Espacio disco duro de 250 Gb en adelante

Necesidades de software

- ✓ PostgreSQL Versión 11 o superior.
- ✓ IDE: Netbeans 8.0.2 o 8.1.
- ✓ Complemento PGAGENT de postgresQL Versión 4.0.0.1
- ✓ Complemento NpgSQL de PostgreSQL Versión 3.2.6.3.
- ✓ JAVA JDK 8 o superior.
- ✓ WEKA 3.8.4 o superior.

2. INSTALACION DE SOFTWARE

2.1. Instalación de JDK y JRE

Para la instalación del JDK, es necesario descargar de la página de Oracle el siguiente componente:

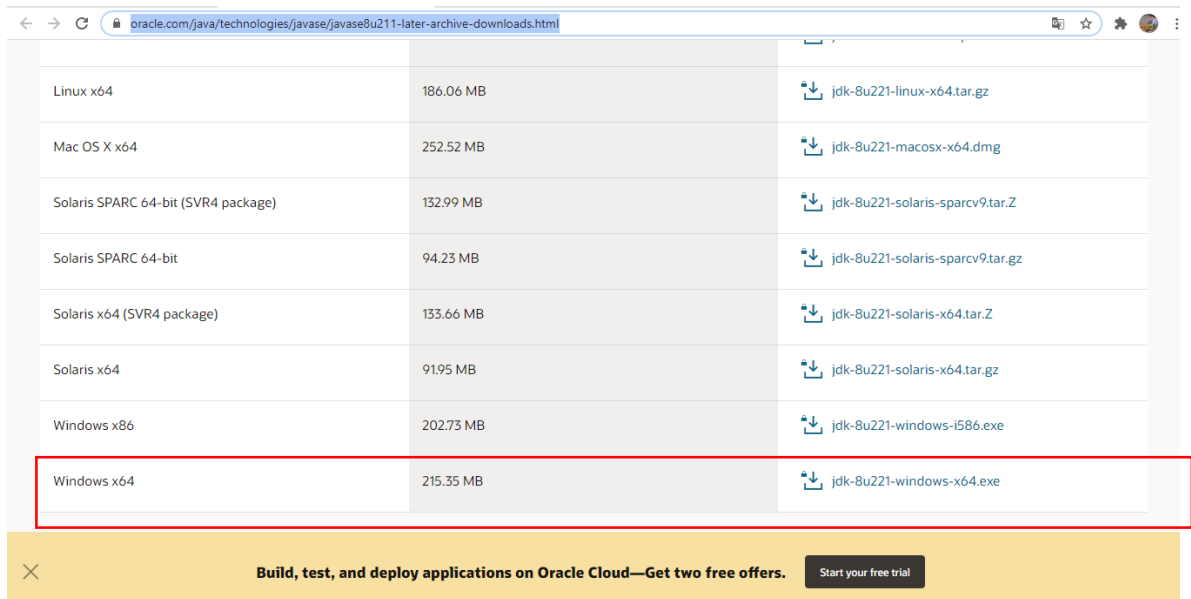


Imagen 1. Página de descarga del JDK

Una vez se ha descargado el archivo .EXE, se procede a la ejecución del mismo

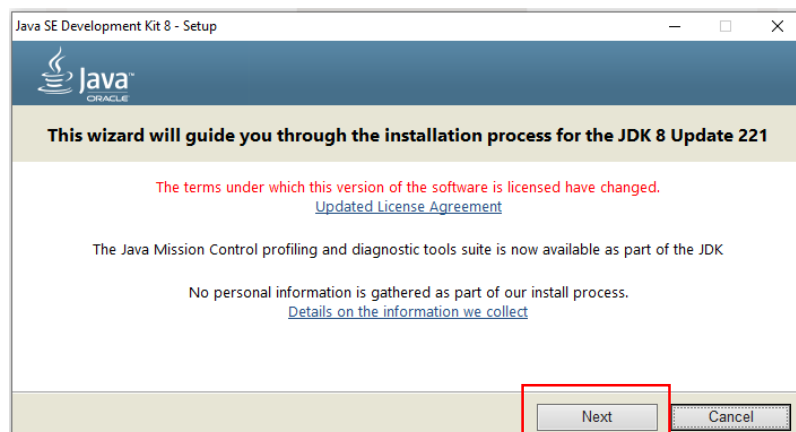


Imagen 2. Ejecución de archivo JDK

Se debe dar clic en el botón “Next”, una vez seleccionado se mostrara una nueva ventana con los componentes a instalar

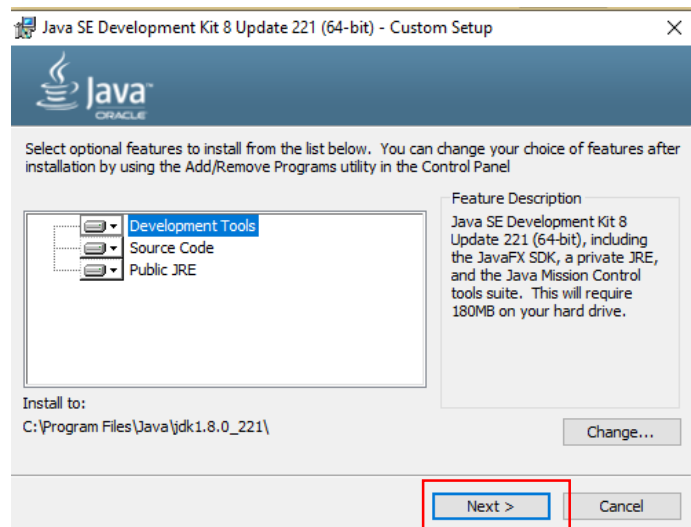


Imagen 3. Ventana de componentes a instalar

Nuevamente se debe seleccionar el botón “Next”, allí aparecerá una barra de carga indicando el estado del proceso de copiado de archivos.

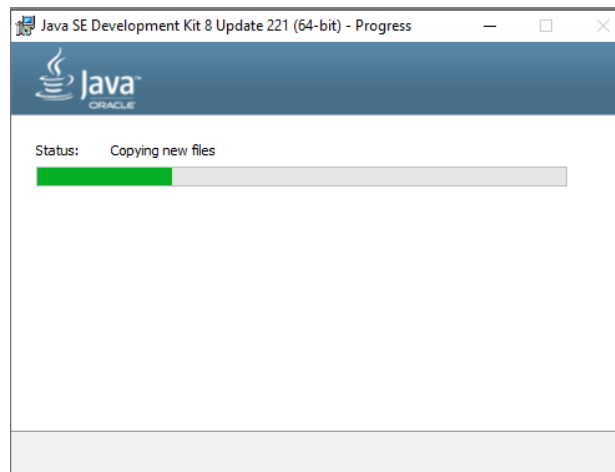


Imagen 4. Proceso de extracción de archivos

A continuación se abrirá una nueva ventana de configuración, allí se debe especificar en qué lugar quedara guardado el JDK.

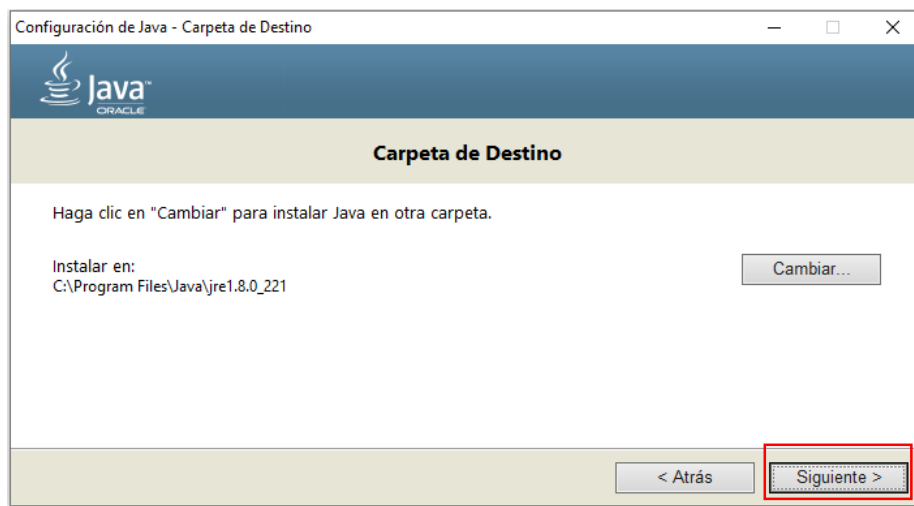


Imagen 5. Configuración de la carpeta de destino

Posteriormente seleccionaremos el botón siguiente, allí iniciara el proceso de instalación de los componentes.



Imagen 6. Instalación de componentes

Una vez ha terminado el proceso cerramos la ventana.

Para la configuración de la variable de entorno de JAVA se debe buscar dentro del panel de control la opción propiedades del sistema, una vez ingresemos a esa ventana seleccionaremos la opción variables de entorno.

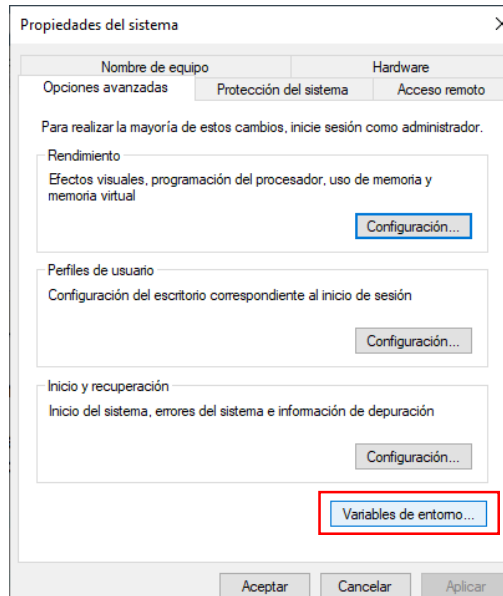


Imagen 7. Ventana propiedades del sistema

Dentro de la ventana de variables de entorno, seleccionaremos la opción de “Nueva...”

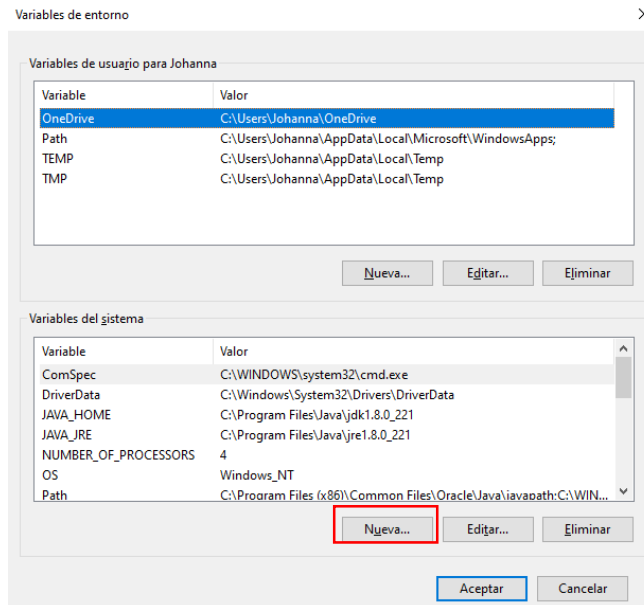


Imagen 8. Ventana variables de entorno

Luego se debe diligenciar el siguiente formulario, en el campo nombre se debe indicar la siguiente información “JAVA_HOME” y en el valor se debe poner la ruta de ubicación del JDK.

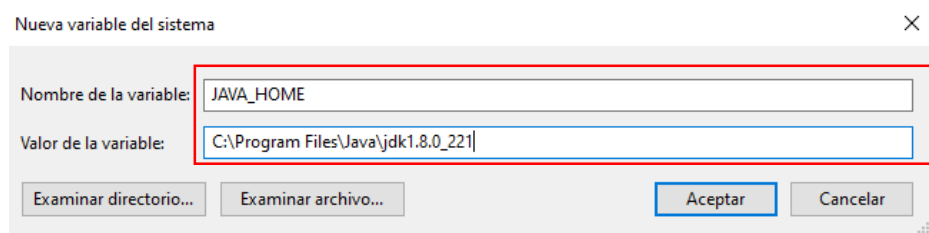


Imagen 9. Configuración inicial de la variable de entorno

Ahora se debe editar la variable PATH, allí se debe adicionar la variable que se ha creado anteriormente

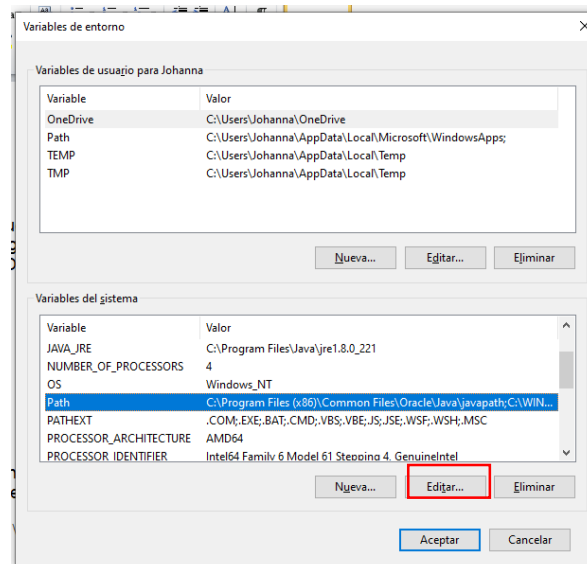


Imagen 10. Configuración variable PATH

Una vez habilitado el campo se debe adicionar al final del texto la siguiente información “C:\WINDOWS;C:\WINDOWS\system32;%JAVA_HOME%\bin”, a continuación le damos aceptar.

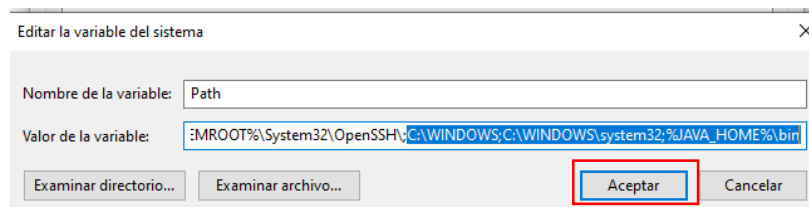


Imagen 11. Edición variable PATH

2.2. Instalación de Netbeans 8.1

Para la instalación se debe descargar de la página de oficial de Netbeans la siguiente herramienta para Java EE.

NetBeans IDE 8.0.2 Download [8.0.1](#) | [8.0.2](#) | [Development](#) | [Archive](#)

Email address (optional):

Subscribe to newsletters: Monthly Weekly
NetBeans can contact me at this address

IDE Language: Platform:

Note: Greyed out technologies are not supported for this platform.

NetBeans IDE Download Bundles

Supported technologies *	Java SE	Java EE	C/C++	HTML5 & PHP	All
<input checked="" type="checkbox"/> NetBeans Platform SDK	•	•			•
<input checked="" type="checkbox"/> Java SE	•	•			•
<input checked="" type="checkbox"/> Java FX	•	•			•
<input checked="" type="checkbox"/> Java EE		•			•
<input checked="" type="checkbox"/> Java ME					•
<input checked="" type="checkbox"/> HTML5		•		•	•
<input checked="" type="checkbox"/> Java Card™ 3 Connected					•
<input checked="" type="checkbox"/> C/C++			•		•
<input checked="" type="checkbox"/> Groovy					•
<input checked="" type="checkbox"/> PHP				•	•
Bundled servers					
<input checked="" type="checkbox"/> GlassFish Server Open Source Edition 4.1		•			•
<input checked="" type="checkbox"/> Apache Tomcat 8.0.15		•			•

Download buttons: Java SE (Free), Java EE (Free), C/C++ (Free), HTML5 & PHP (Free), All (Free)

Imagen 12. Descarga de Netbeans

Una vez se ha descargado el archivo ejecutable, procedemos a iniciar la instalación. En la primera parte esperamos a que se configuren algunos archivos y se descarguen los componentes.

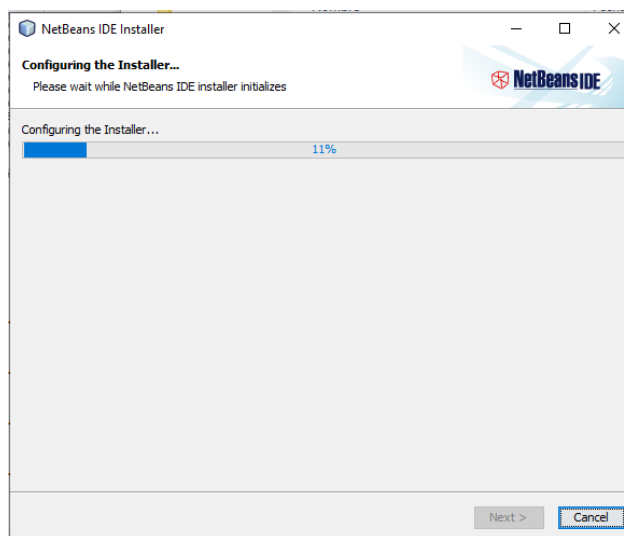


Imagen 13. Configuración de componentes

Una vez se ha descargado y configurado la instalación, aparecerá una vista indicando que servidores se desean descargar para nuestro caso solo sería seleccionar GlassFish.

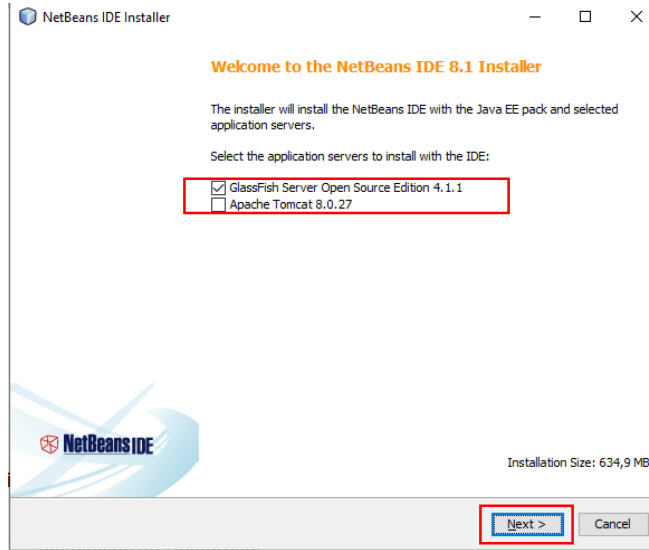


Imagen 14. Selección del servidor de despliegue

Al dar clic en "Next", se deben aceptar los términos y condiciones de licencia

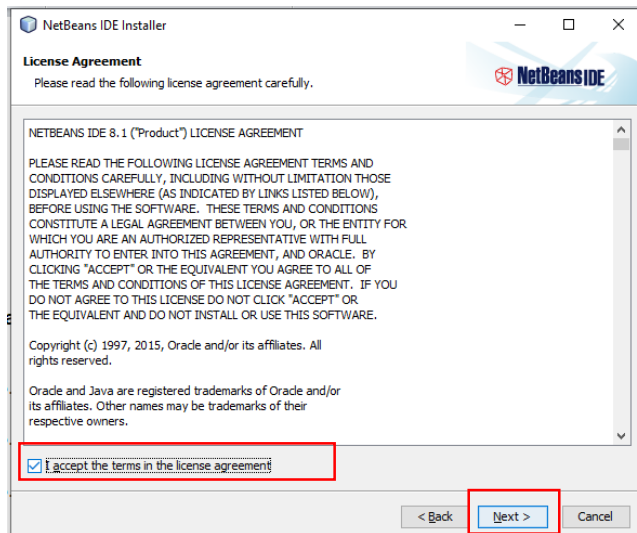


Imagen 15. Aceptación de términos y condiciones

Luego de seleccionar “Next”, aparecerá un formulario en el cual se debe indicar el folder de instalación y la ubicación del JDK.

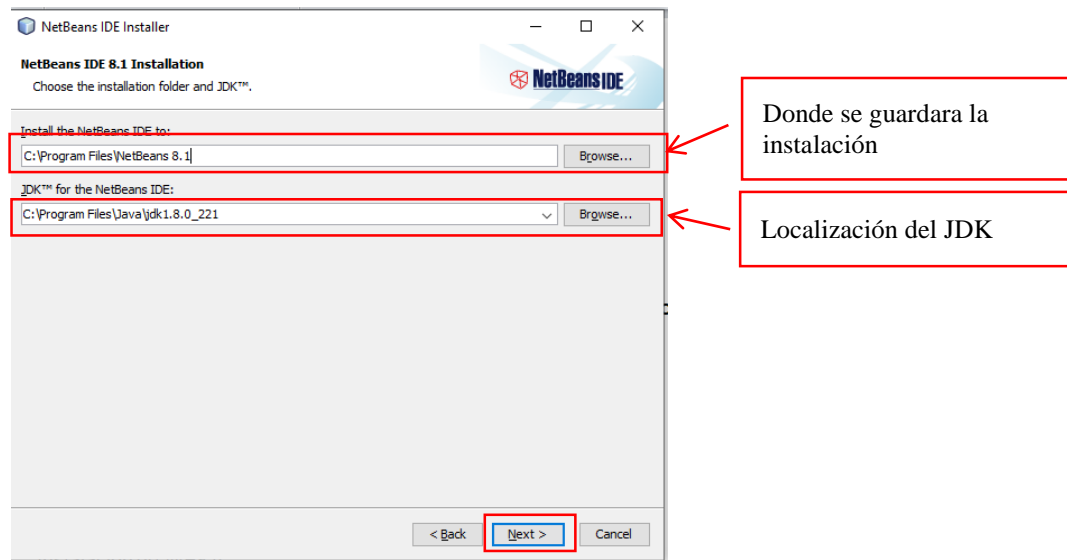


Imagen 16. Configuración y localización de carpetas

A continuación se mostrara un detallado de la localización y el peso del aplicativo.

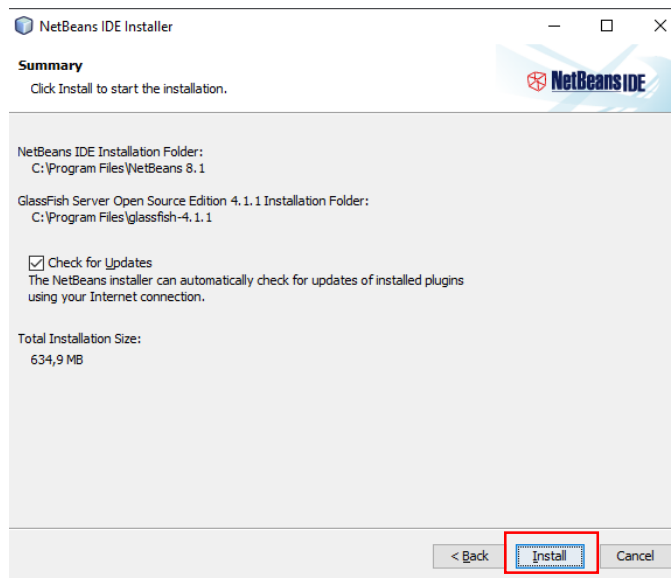


Imagen 17. Información detallada de la instalación del Netbeans.

Procedemos a seleccionar el botón instalar. Aproximadamente la instalación tardara entre 10 a 15 minutos.

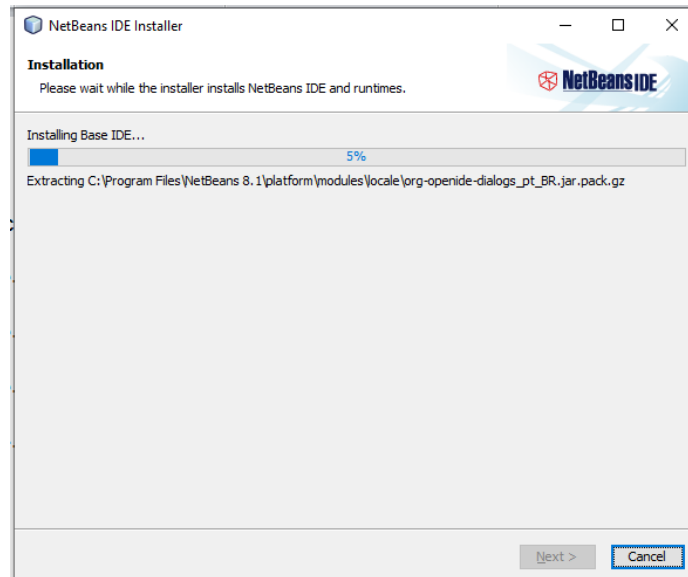


Imagen 18. Instalación de Netbeans

Al finalizar la instalación aparecerá unas indicaciones sobre el manejo de Netbeans.

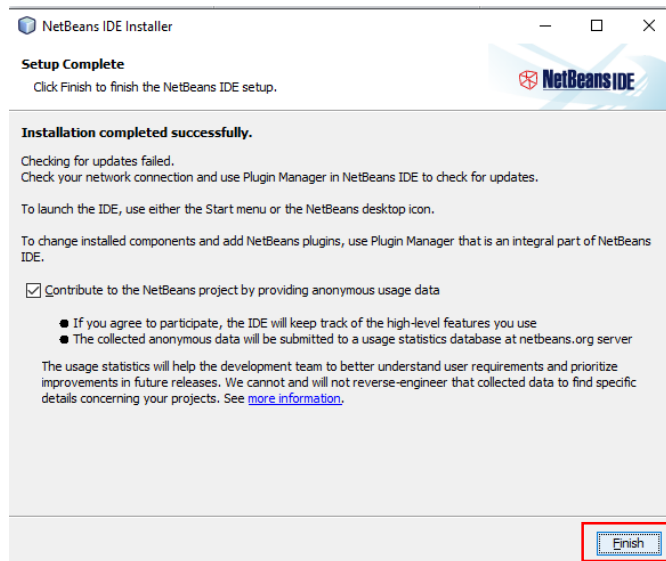


Imagen 19. Instalación completada de Netbeans

A continuación se muestra el entorno grafico de Netbeans

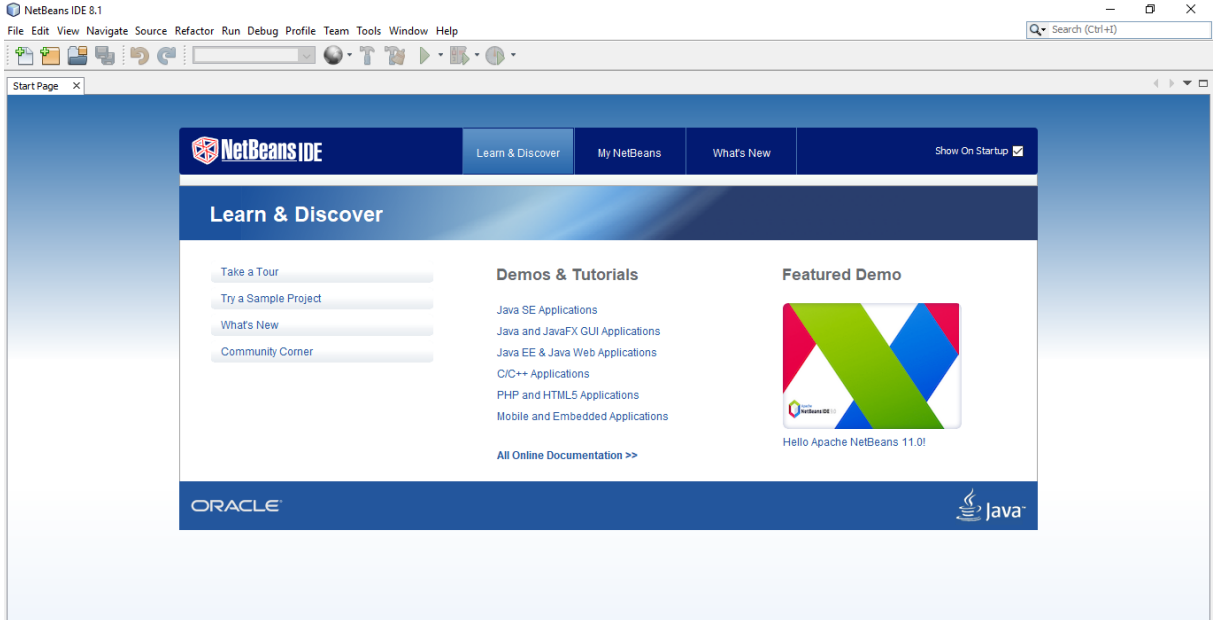


Imagen 20. Entorno de Netbeans

Para añadir el servidor primero seleccionaremos del menú principal la opción Tools, luego la opción Servers

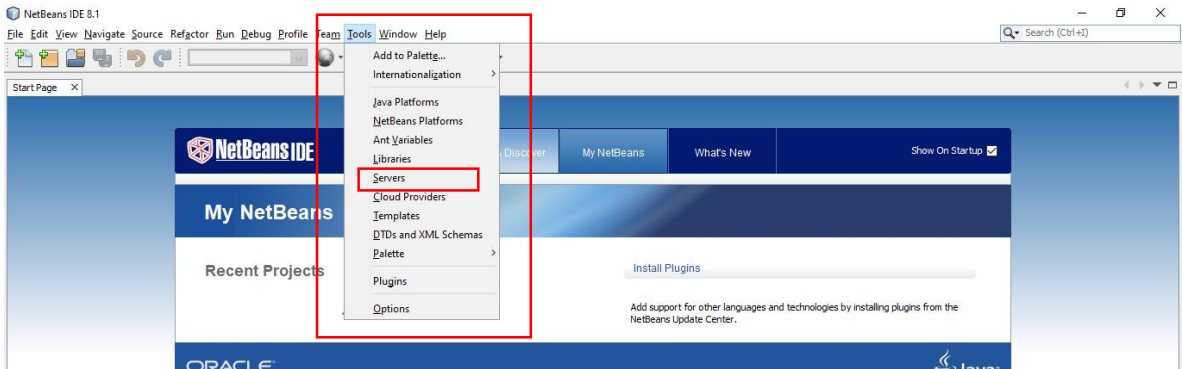


Imagen 21. Configuración de servidor de despliegue

A continuación seleccionaremos la opción “Add Server”



Imagen 22. Añadir servidor

Luego seleccionamos Glassfish Server, y configuraremos la ruta del servidor de la siguiente manera:

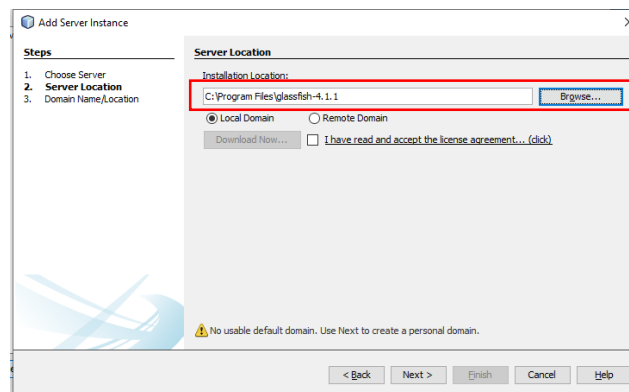


Imagen 23. Configuración de ruta de GlassFish

Una vez seleccionado el botón “Next”, verificaremos que los datos sean correctos.

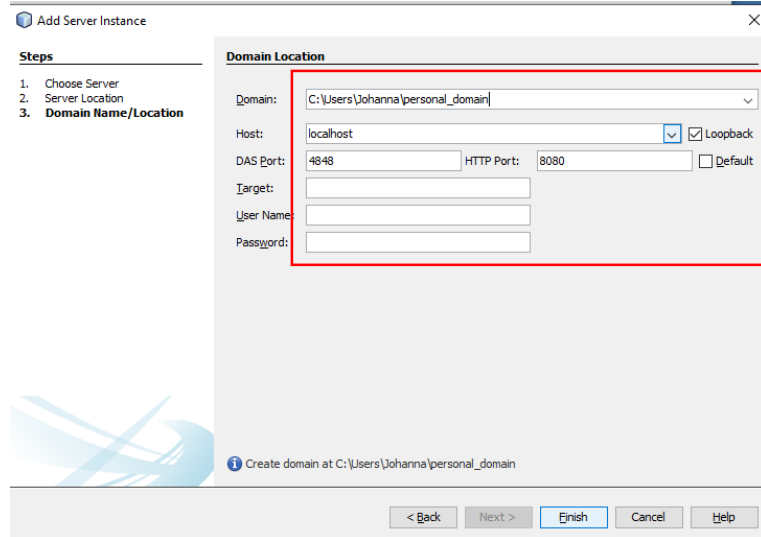


Imagen 24. Configuración del host y el puerto

Por ultimo damos clic sobre el botón “Finish”, al final aparecerá en la consola el estado de algunas configuraciones del servidor.

```
Output - Output x
Running the create-domain subcommand
Using port 32228 for Admin.
Using port 32260 for HTTP Instance.
Using port 32256 for JMS.
Using port 32217 for IIOP.
Using port 32261 for HTTP_SSL.
Using port 32219 for IIOP_SSL.
Using port 32219 for IIOP_MUTUALAUTH.
Using port 32266 for JMX_ADMIN.
Using port 32246 for OSGI_SHELL.
Using port 32189 for JAVA_DEBUGGER.
Distinguished Name of the self-signed X.509 Server Certificate is:
[CN=DESKTOP-443HDAV,OU=GlassFish,O=Oracle Corporation,L=Santa Clara,ST=California,C=US]
Distinguished Name of the self-signed X.509 Server Certificate is:
[CN=DESKTOP-443HDAV-instance,OU=GlassFish,O=Oracle Corporation,L=Santa Clara,ST=California,C=US]
Domain personal_domain created.
Domain personal_domain admin port is 32228.
Domain personal_domain allows admin login as user "anonymous" with no password.
Command create-domain executed successfully.
```

Imagen 25. Instalación de GlassFish finalizada.

2.3. Instalación de PostgreSQL

Para la instalación se debe ingresar a la página de EDB y allí seleccionar el siguiente componente:

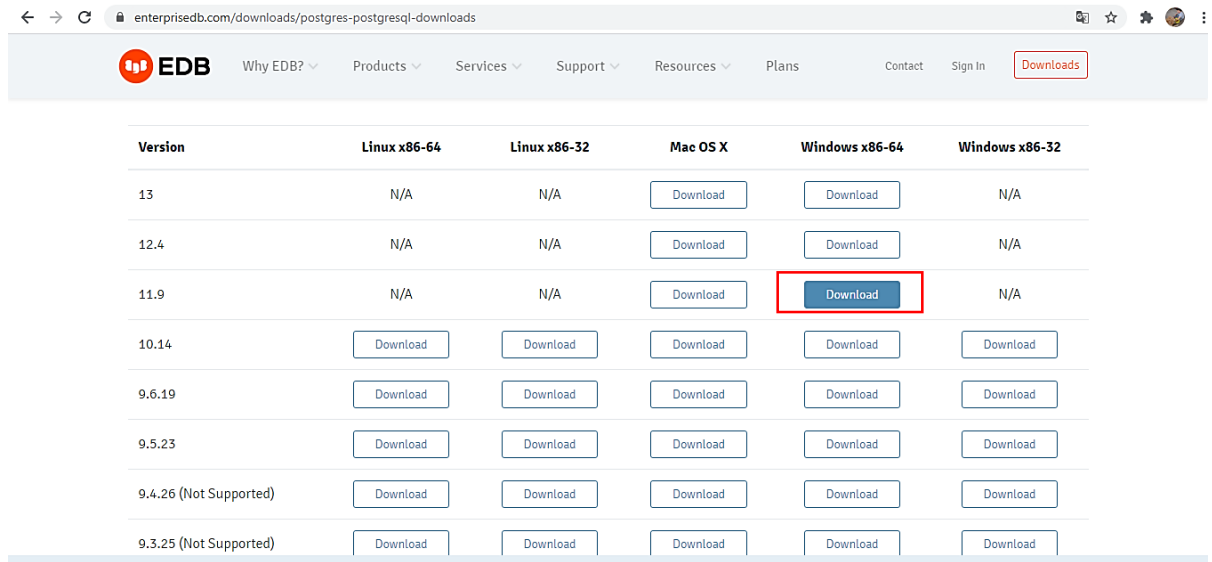


Imagen 26. Descarga del PostgreSQL

Una vez descargado el ejecutable procedemos a iniciar la instalación.

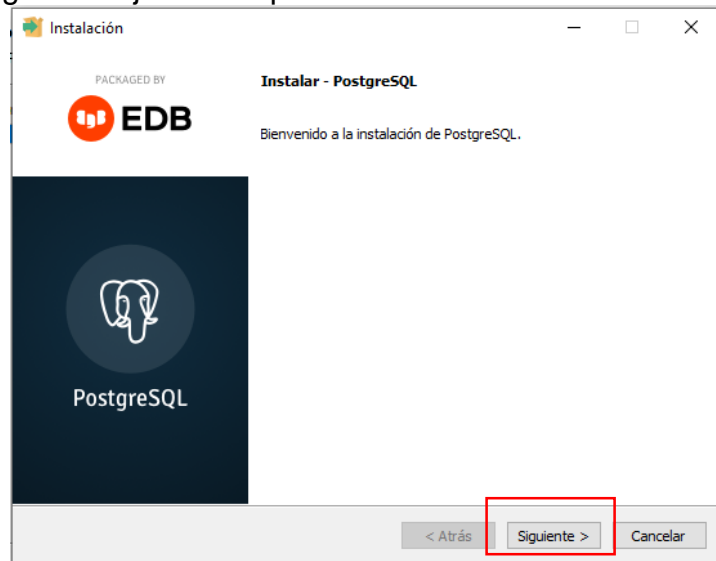


Imagen 27. Ventana principal de instalación del PostgreSQL

A continuación se debe especificar el directorio donde quedara guardado el programa.

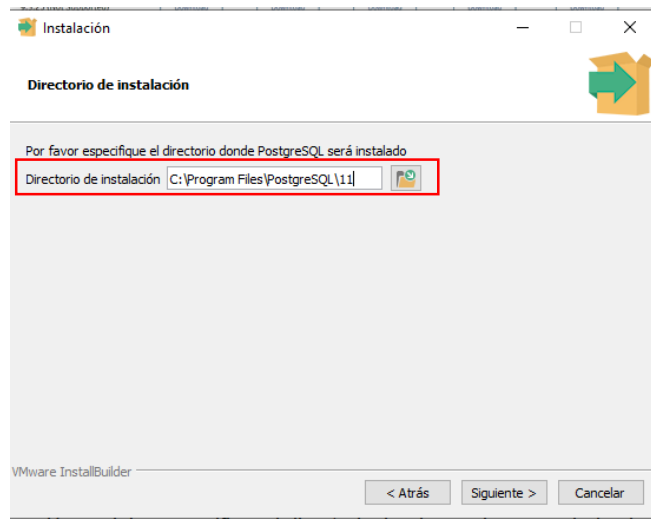


Imagen 28. Configuración de directorio de PostgreSQL

Luego seleccionaremos que componentes de PostgreSQL queremos instalar, para nuestro caso dejaremos seleccionado como esta por defecto. Damos clic en siguiente.

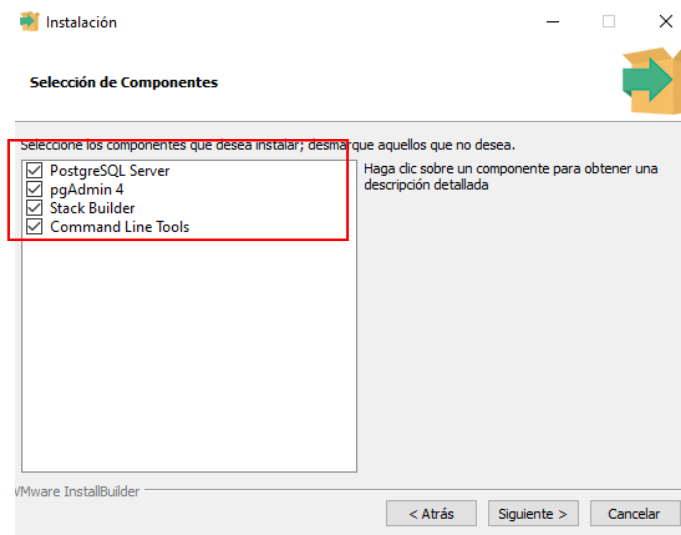


Imagen 29. Configuración de componentes a instalar

Ahora aparecerá un formulario en el cual debemos indicar la contraseña de acceso a PostgreSQL.

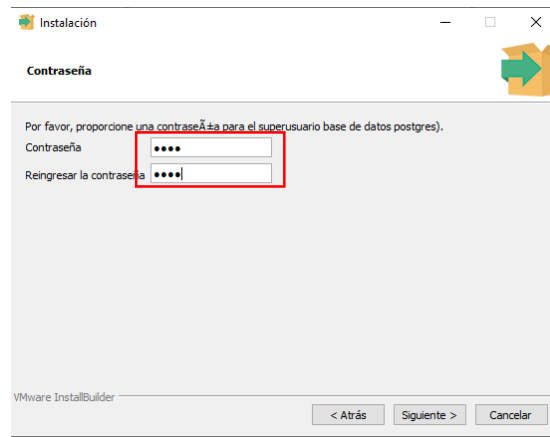


Imagen 30. Configuración de contraseña

A continuación solicitara ingresar el puerto de conexión con el servidor, por defecto siempre es el 5432.

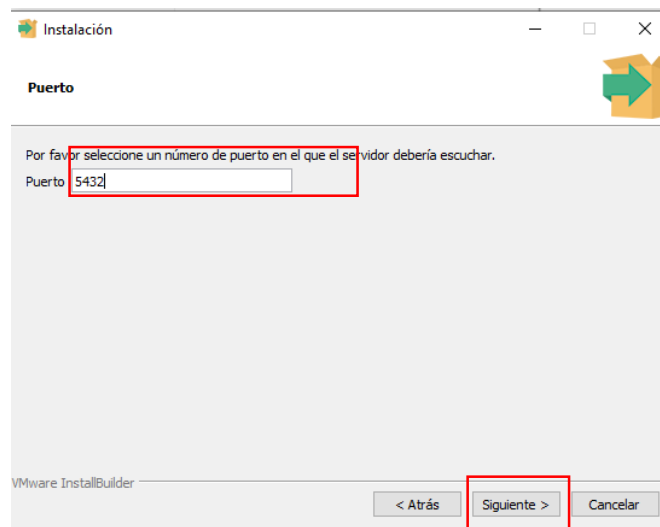


Imagen 31. Configuración del puerto de conexión

Luego de dar clic en siguiente, se debe seleccionar la configuración regional, por defecto dejamos la que está.

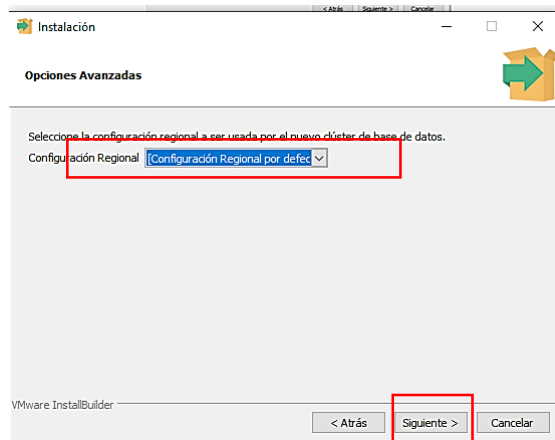


Imagen 32. Configuración regional

Por ultimo esperamos a que termine la instalación.

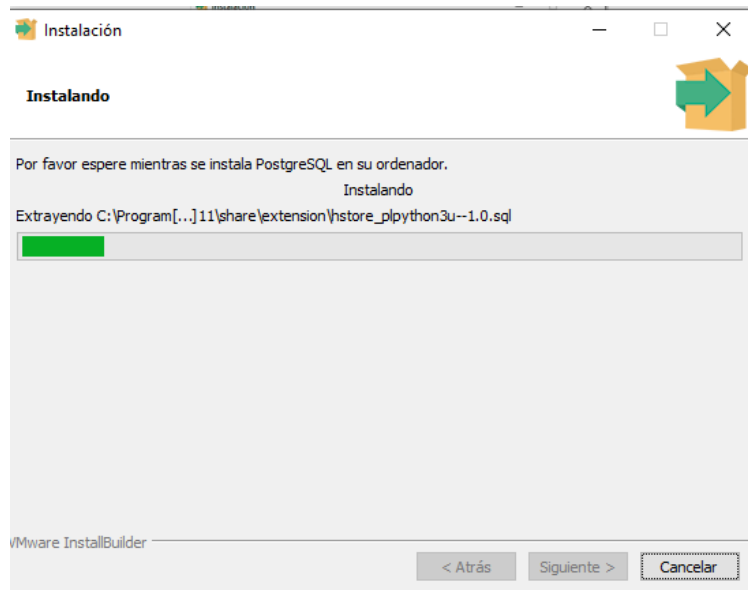


Imagen 33. Instalación de postgresQL

2.3.1. Instalación de extensiones Npgsql y PgAgent

Para instalar estas extensiones debemos ejecutar el programa Stack Builder 4.2.0, una vez iniciada la herramienta, seleccionaremos la opción “PostgreSQL 11”.

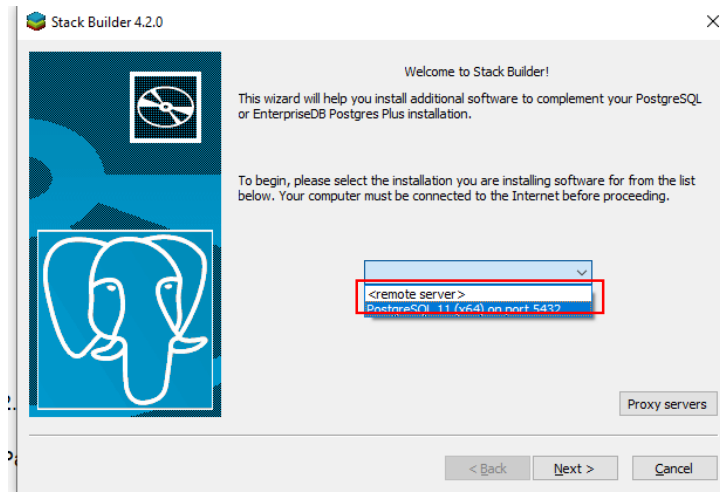


Imagen 34. Selección del gestor de base de datos

A continuación seleccionaremos las extensiones que queremos instalar

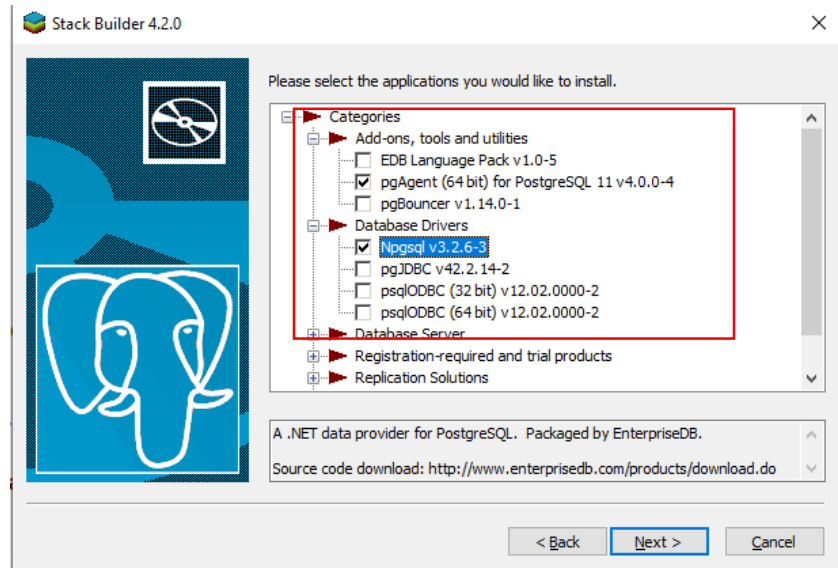


Imagen 35. Selección de las extensiones a instalar

Una vez seleccionado el botón “Next” seleccionaremos el directorio donde quedaran guardadas las extensiones.

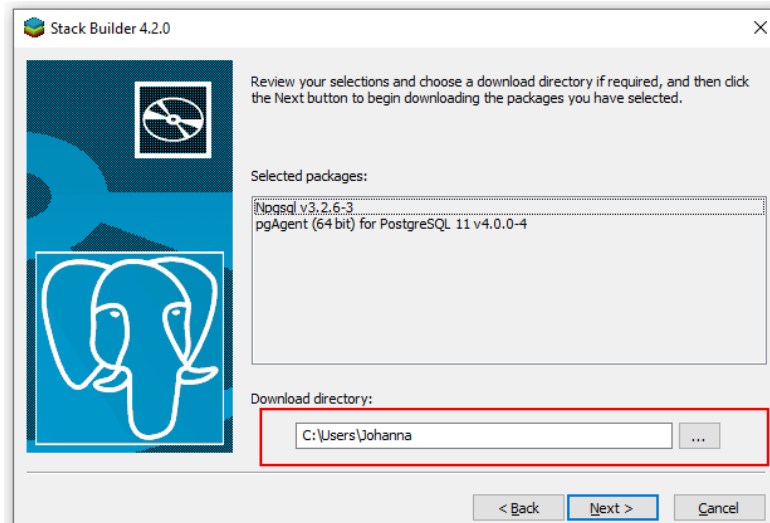


Imagen 36. Configuración del directorio de descarga

Esperamos a que se descarguen las extensiones, una vez termine aparecerá una ventana nueva y allí daremos clic en el botón “Next” hasta llegar a esta vista:

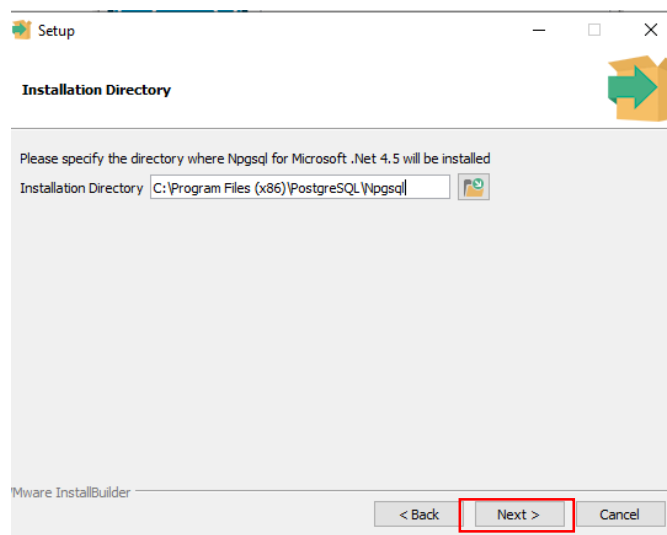


Imagen 37. Configuración del directorio de instalación de extensiones

Por cada extensión descargada aparecerá dicha ventana, allí seleccionaremos el lugar de instalación de la extensión. Posteriormente, damos clic al botón “Next” y esperamos a que se instale la extensión. Estos últimos pasos se repetirán de acuerdo a la cantidad de extensiones descargadas

2.4. Instalación de WEKA

Para la descarga de WEKA debemos ingresar a la página de SourceForge y buscar la herramienta WEKA.

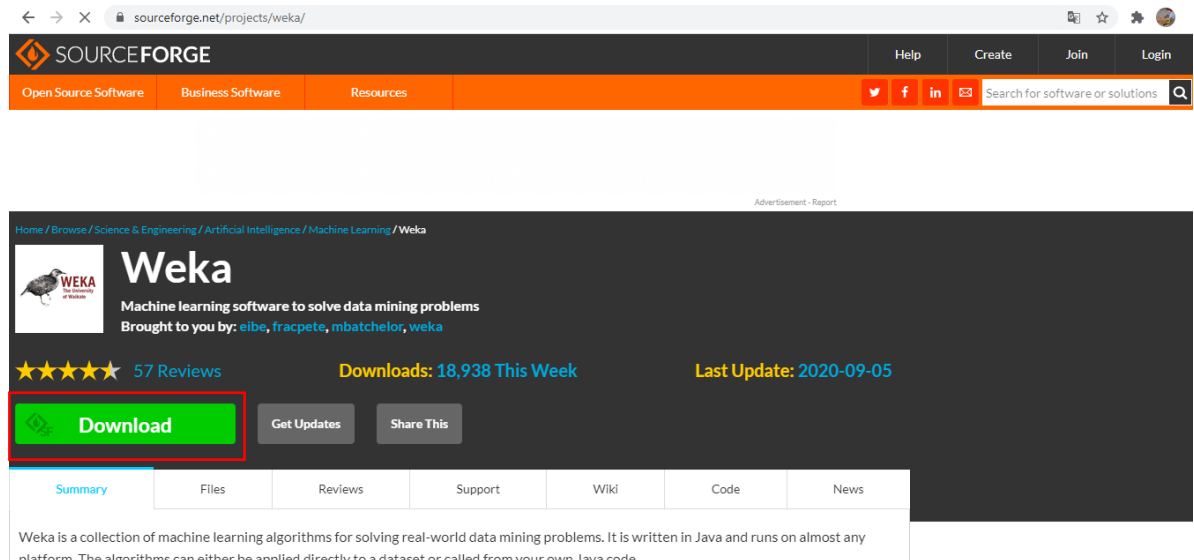


Imagen 38. Descarga de WEKA

Procedemos a instalar el programa, en la primera parte debemos aceptar términos de licencia para ello seleccionaremos el botón “I Agree”.

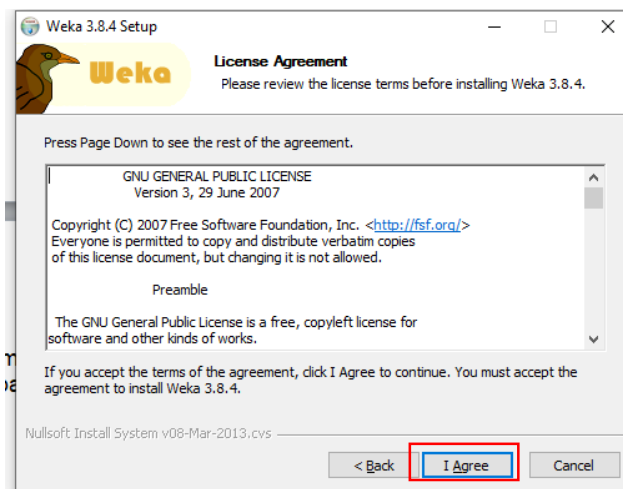


Imagen 39. Términos y condiciones de licencia

Posteriormente seleccionaremos los componentes a instalar, por defecto dejamos seleccionada la opción Full

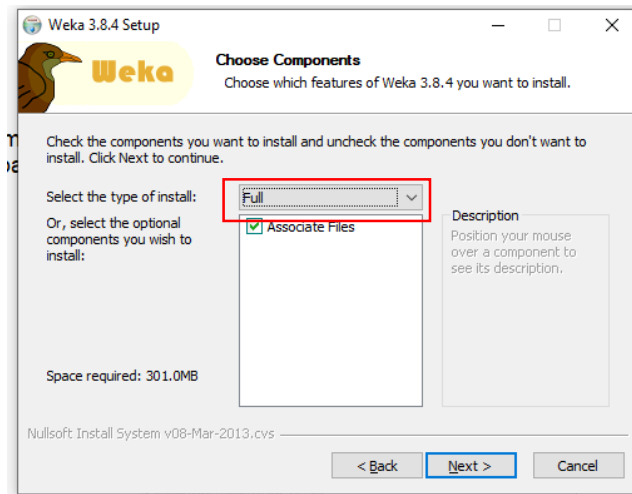


Imagen 40. Configuración en la instalación de componentes

Una vez seleccionado el botón "Next", debemos indicar en donde queremos guardar la instalación del programa.

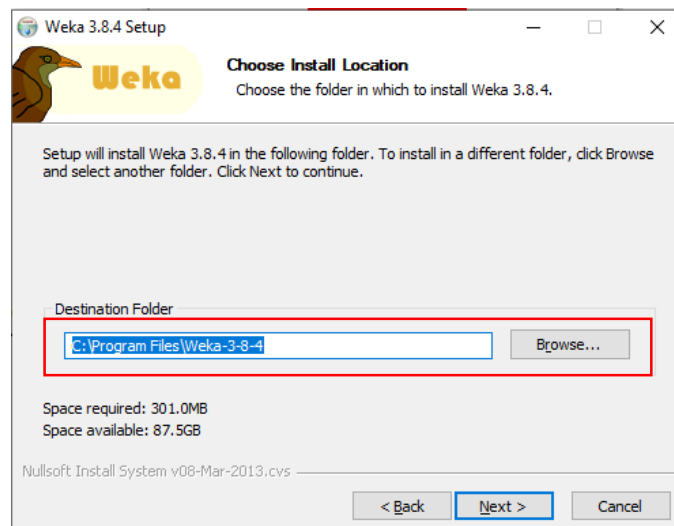


Imagen 41. Configuración del directorio

Igualmente debemos seleccionar el folder de inicio, esta configuración se deja por defecto.

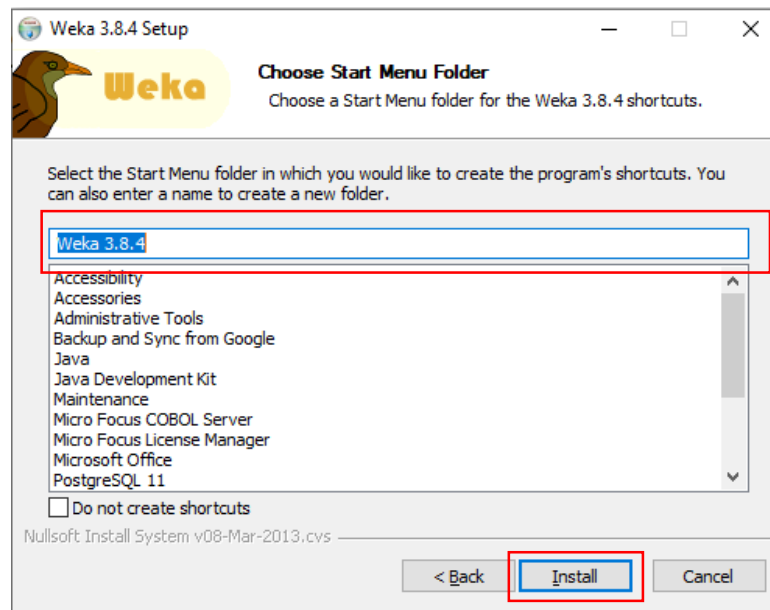


Imagen 42. Configuración del folder inicial

Por ultimo damos clic sobre el botón "install", esperamos a que instale todos los componentes.

ANEXO C Artículo CACIED

Análisis descriptivo y predictivo de brechas de género en las pruebas Saber

Descriptive and predictive analysis of gender gaps in Saber tests

Alisson Catalina Celeita Gutiérrez  <https://orcid.org/0000-0001-9750-2848>

Universidad de Cundinamarca. Facatativá, (Colombia)

acceleita@ucundinamarca.edu.co

Angie Paola Manrique Ravelo  <https://orcid.org/0000-0003-0302-406X>

Universidad de Cundinamarca. Facatativá, (Colombia)

apaolamanrique@ucundinamarca.edu.co

Corin Jazmín Viracacha Viracacha  <https://orcid.org/0000-0003-4487-938X>

Universidad de Cundinamarca. Facatativá, (Colombia)

cviracacha@ucundinamarca.edu.co

Cesar Yesid Barahona Rodríguez  <https://orcid.org/0000-0001-7673-7381>

Universidad de Cundinamarca. Facatativá, (Colombia)

cbarahona@ucundinamarca.edu.co

Resumen

A través del tiempo, en Colombia, se han presentado brechas de género en gran parte de los campos económicos y socioculturales. Asimismo, centrándose en el ámbito educativo, hay estudios que demuestran que allí es donde comienza esta diferencia. Por esta razón, es importante realizar un análisis de los resultados de las pruebas del estado desde el año 2016 y de esta forma hallar los factores que más inciden en esta problemática.

El objetivo principal de este proyecto es analizar los datos brindados por el ICFES de las pruebas Saber para encontrar indicadores de brechas de género y poder relacionar los núcleos temáticos en común desde las pruebas presentadas en la educación básica hasta la educación superior y lograr predecir estos comportamientos. El análisis de la información se realizará por medio de técnicas de minería de datos, tanto descriptivas como predictivas, por su parte, se hará uso de algoritmos matemáticos para calcular la brecha de género. Gracias a esto, se le presentará al usuario una representación detallada de los indicadores hallados, con el fin de que sean usados en futuras investigaciones de entes que se interesen en el estudio y la reducción de esta problemática, las cuales contribuyan en la toma de decisiones que impulsen el desarrollo país. Concluyendo, para brindar la información de una forma más clara se requiere el desarrollo de un aplicativo web.

Abstract

Over time, in Colombia, gender gaps have occurred in much of the economic and socio-cultural fields. Also, focusing on the educational field, there are studies that show that this is where

this difference begins. For this reason, it is important to carry out an analysis of the results of the state tests since 2016 and in this way find the factors that most affect this problem.

The main objective of this project is to analyze the data provided by the ICFES of the Saber tests to find indicators of gender gaps and to be able to relate the thematic nuclei in common from the tests presented in basic education to higher education and to predict these behaviors. The analysis of the information will be carried out through data mining techniques, both descriptive and predictive, on the other hand, mathematical algorithms will be used to calculate the gender gap. Thanks to this, the user will be presented with a detailed representation of the indicators found, so that they can be used in future investigations of entities that are interested in the study and the reduction of this problem, which contribute to decision-making, that promote the development of the country. In conclusion, to provide the information in a clearer way, the development of a web application is required.

Palabras clave

Minería de datos; Brechas de género; Educación; Big Data; Análisis descriptivo, Análisis predictivo.

I. INTRODUCCION

Durante mucho tiempo se han presentado brechas de género en gran parte de los campos económicos y socioculturales del mundo. Al observar cada uno de los ámbitos se puede concluir que es una problemática que se encuentra muy lejos de ser solucionada. En el ámbito

laboral se aprecia como las ganancias del sexo femenino son 20.2% menores respecto a las generadas por los hombres, a pesar de que está comprobado que ellas son las que más horas trabajan [1]. La situación no cambia con los cargos gerenciales, para estos las mujeres ocupan únicamente un 16.2% según el estudio “Report on Women Board Directors of Latin America’s 100 Largest Companies” del año 2016, debido a la existencia de estereotipos de dependencia y limitaciones a la hora de conseguir empleo. Esta dificultad como ya se ha mencionado anteriormente no es exclusiva de un área en específico, por ejemplo, de acuerdo con los reportes de la UNESCO en el año 2014, solamente en la educación primaria persistía la disparidad de género en el 37% de los países, así mismo las niñas han conformado el 53% de la población que no recibe algún tipo de educación y se sabe que la mitad de estas nunca asistirán a una escuela [2].

En Colombia según el “Global Gender Gap Report” las brechas de género más significativas se encuentran en el liderazgo político y el fortalecimiento económico. En el caso del consolidamiento político, conforme la resolución sobre la participación de la mujer en la política aprobada por la Asamblea General de las Naciones Unidas en 2011, en todo el mundo, las mujeres siguen siendo marginadas en este campo a causa de diversos factores como lo son los estereotipos, bajos niveles de educación, afectación de manera desproporcionada de la pobreza, etc. Para el año 2014 el puesto del país descendió 12 puestos en la medición en empoderamiento político del Índice de Brechas de Género, ya que a pesar de ser el 52% de la población colombiana, solo el 12% de esta cifra ocupa algún cargo de elección popular [3].

El origen de toda esta problemática por lo general se da desde la primera infancia, puesto que la educación recibida de una u otra manera contribuye a que los hombres se destaquen por encima del género femenino, o en otras palabras, “Desde la escuela no se están propiciando los espacios suficientes para que las mujeres sean líderes y lleguen a donde tradicionalmente han estado subrepresentadas” [4]. Este tipo de deficiencias y algunas similares hacen que el alcance de las brechas de

género incrementa, mostrando así que la calidad educativa empezando por el nivel más básico es indispensable en la disminución de esta inequidad.

En varios países se han realizado estudios que demuestran que al tener un sistema educativo de calidad las diferencias son casi nulas para hombres y mujeres, como es el caso de Estados Unidos que ha venido disminuyendo la inequidad entre los dos sexos. Si en 1970 sólo una de cada 13 personas capaces de obtener 700 puntos en la prueba SAT (Scholarship Aptitude Test) antes de los 13 años era mujer, hoy la cifra es de 1 a 3 y se sigue cerrando camino a la equidad. Otro caso importante para destacar se observa en Marruecos, donde los niños estudian en centros educativos distintos a los de las niñas y con este tipo de enseñanza no se presenta una presión social lo que hace que la brecha de género educativa sea inexistente [5].

Por otra parte, al hacer énfasis en la educación superior en las áreas de la ciencia, tecnología, ingeniería y matemáticas (STEM, siglas en inglés), en los países de América Latina y el Caribe (ALC) se han venido presentando importantes diferencias en el porcentaje de mujeres graduadas de programas educativos pertenecientes a estos núcleos, entre 20% y 50% en el año 2015; mientras que en otras zonas de estudio como lo son las ciencias naturales y afines el porcentaje es mucho mayor encontrándose en un rango del 50% y 80%. A continuación, se muestra en la Fig. 1 que las mujeres profesionales en las tecnologías de la información y la comunicación (TIC) o en ingeniería, industria y construcción es significativamente más bajo que el encontrado en ciencias naturales, matemáticas y estadística. Cabe aclarar que los datos no demuestran los diferentes niveles de estudios, como los son por ejemplo maestrías, especializaciones y/o doctorados, pero sí refleja una conducta generalizada en ALC donde las mujeres no presentan el mismo comportamiento en los diversos campos del conocimiento STEM [6].

En el país para el año 2012, investigadores de la Universidad Tecnológica de Bolívar emplearon técnicas de contraste de medias (t Student) y regresiones por

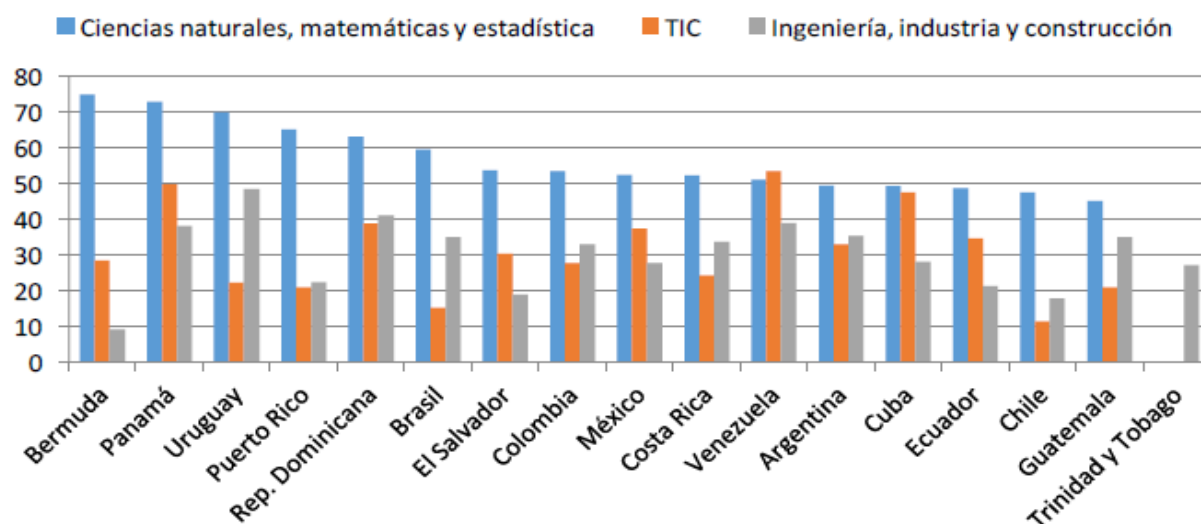


Fig. 1 Porcentaje de mujeres entre los graduados de programas de educación en campos STEM. Fuente: López-Bassols, V., Grazi, M., Guillard, C., & Salazar, M. [6]

cuantiles en los resultados de las pruebas Saber 11 del año 2009, haciendo énfasis en los núcleos de matemáticas y lenguaje; con los resultados obtenidos se pudo concluir que en los componente de matemáticas los hombres tienen un mejor desempeño que las [7]. Son diversos los factores que pueden afectar el crecimiento o la reducción de la brecha de género, se busca hacer un análisis tanto descriptivo como predictivo en las pruebas estandarizadas nacionales de todos los niveles, desde la educación básica hasta la educación técnica, tecnológica y superior; con el fin de determinar estos indicadores de brecha de género.

En la actualidad, lo más usado para realizar análisis de información es la minería de datos por su capacidad de predicción al ser “un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos [8]. En este proyecto se hará uso de algunos algoritmos de Data Mining como mecanismo de estudio y evaluación de la información de las pruebas, no sin antes hacer un análisis del comportamiento de los mismos para así conocer cuál o cuáles presentan mejores resultados.

El análisis predictivo se define como el proceso de descubrir patrones significativos de datos utilizando técnicas de reconocimiento, estadísticas, aprendizaje automático, inteligencia artificial y minería de datos [9]. Es todo un proceso analítico que se desarrolla de forma automática o sistematizada con el fin de descubrir datos futuros a partir de datos pasados. La selección del modelo más conveniente para el estudio se hace en relación de los datos disponibles, en términos generales cualquier algoritmo de predicción se encarga de encontrar patrones y relaciones ya sea de variables dependientes y/o independientes; la efectividad de los mismos se determina a través de coeficientes de relación de regresión donde los errores deben ser mínimos para que el modelo sea considerado como óptimo.

Existen 2 tipos de modelos predictivos, los de clasificación y de regresión. Los primeros se encargan de predecir la pertenencia a una clase específica, teniendo como resultado binarios, o un sí o un no (en forma de 0 y 1) con su respectivo grado de probabilidad; mientras que los segundos permiten la predicción de un valor, por ejemplo, para el caso de estudio actual son prácticos para hallar porcentajes de relación en cada uno de los indicadores de brecha de género y entre núcleos temáticos de las diferentes pruebas. Por otra parte, las técnicas más usadas son los árboles de decisión, que son modelos de clasificación que dividen los datos en subconjuntos basados en categorías de variables de entrada; las técnicas de regresión lineal y logística siendo uno de los métodos más empleados en estadística por medio de las cuales se estiman las relaciones entre variables, y por último, las redes neuronales que consisten en técnicas sofisticadas capaces de modelar relaciones extremadamente complejas [10].

La metodología de desarrollo del proyecto consta de 3 grandes etapas, la primera de ellas es el análisis descriptivo de cada una de las pruebas Saber, la segunda es el análisis predictivo de las mismas relacionando los núcleos temáticos comunes entre sí y por último el desarrollo de un aplicativo web que permita mostrar la información hallada en las dos etapas anteriores.

Etapas 1: Para el análisis de los resultados de las pruebas Saber, en primer lugar, se debe hacer uso de los datos abiertos proporcionados por el ICFES (Instituto Colombiano para la Evaluación de la Educación) en su repositorio de acceso abierto, siendo esto un gran volumen de información, a la hora de ser examinada se vuelve en una tarea muy compleja, por esta razón se hará uso de la minería de datos educativa puesto que facilita el procesamiento de la misma. Además, tiene como ventaja con respecto a los métodos estadísticos, el descubrimiento de información oculta y el manejo óptimo de los datos, convirtiéndose en una herramienta más eficaz al momento de la investigación.

Como primer opción de técnica de análisis se ha optado por usar el algoritmo K-Medias, sin embargo, antes de iniciar este proceso descriptivo se ha decidido realizar un estudio del comportamiento de varios algoritmos de minería de datos para poder escoger el más adecuado. A partir de este momento, el desarrollo del proyecto implementará una de las metodologías más reconocidas en Data Mining, como lo es CRISP-DM, que es considerada la “metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico e industrial” [11], la cual incorpora a su vez la gestión de proyectos y permite un movimiento bidireccional entre sus etapas, a continuación, se muestra los pasos a seguir en esta fase:

- 1) **Comprensión del negocio:** Se pretende desarrollar un modelo descriptivo de los indicadores de brecha de género, el cual relacione aspectos económicos, socioculturales, familiares y personales. Condicionados por los datos disponibles de cada una de las pruebas se ha decidido empezar el análisis desde el periodo académico 2016 hasta su última actualización.

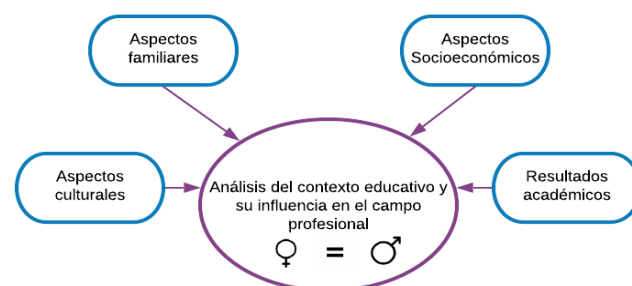


Fig. 2 Análisis del negocio. Fuente: Autor(es)

II. DESARROLLO

- 2) Comprensión de los datos: Recopilación de la información de los repositorios del ICFES. En la cual se obtiene los puntajes globales y específicos presentados por los estudiantes, factores socioeconómicos tanto familiares como personales, localización del estudiante y los datos relacionados con la institución (en unas pruebas sería el tipo, mientras que en otras el nombre de la misma).

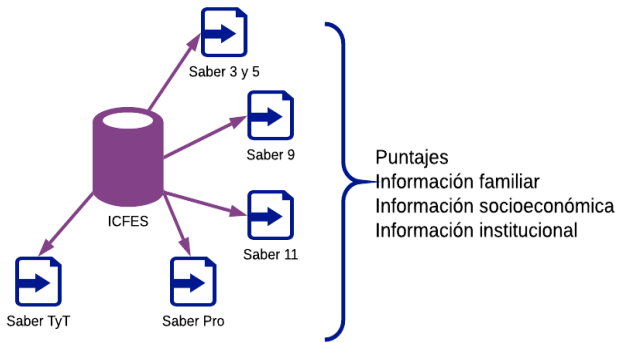


Fig. 3 Recopilación de resultados. Fuente: Autor(es)

- 3) Preparación de los datos: Se empezará en este punto de la investigación el tratamiento de los datos. Se deberá extraer, transformar y cargar la información de manera que quede seleccionada para su procesamiento.



Fig. 4 ETL (Extracción, transformación y carga). Fuente: Autor(es)

- 4) Modelado: En esta fase se hará uso del algoritmo K-Medias como regla de agrupación para caracterizar a los estudiantes según su comportamiento y similitudes encontradas.

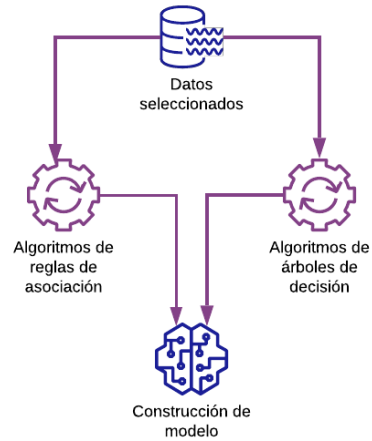


Fig. 5 Creación del modelo con algoritmos de minería de datos. Fuente: Autor(es)

- 5) Evaluar: Se evaluará la calidad del modelo desarrollado, determinando su margen de error y exactitud.

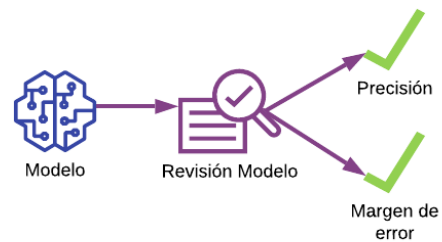


Fig. 6 Verificación del modelo. Fuente: Autor(es)

¿Cuál es el procedimiento para el análisis de conglomerados K-Medias?

Para entrar en contexto, el análisis de conglomerados (cluster analysis) es una técnica multivariante que permite agrupar los casos o variables de un archivo de datos en función del parecido o similitud existente entre ellos [12]. Esta técnica se considera propicia para la investigación porque permite encontrar variables a partir de las características de la información. El uso de la misma inicia con la definición del conjunto de variables que van a ser tratadas, posteriormente se realiza el agrupamiento de los datos. Si la distancia entre los elementos del mismo clúster es pequeña (intra-cluster), indicaría que la similitud para estos casos es alta, de manera que para los elementos externos la similitud es baja. De forma resumida, se puede decir que la fase final es la verificación de convergencia, siendo esta la más importante puesto que describe los resultados obtenidos durante todo el proceso y si alguno de ellos no resulta convergente entonces se deberá volver a iniciar al paso inicial.

Una vez culminado el proceso descriptivo se hará uso de esta información para generar tablas y gráficas en el cual se evidencien los resultados analíticos de las pruebas, las variables halladas y los años de estudio; de manera clara y concisa para el usuario final.

Etapa 2: Posteriormente, se realizará un análisis predictivo para determinar las brechas de género en todos los niveles educativos, para ello se hará uso de la técnica de árboles de decisión, siendo esta una de las más potentes a la hora de analizar decisiones y probabilidades, establecer normas para predecir comportamientos e identificar relación entre las variables halladas. Dentro de esta técnica se busca seleccionar de un conjunto de resultados los que mejor precisión tengan sobre los datos seleccionados.

Con la determinación de las predicciones se quiere mostrar la relación, si es que existe, entre los núcleos temáticos en común, visualizando su comportamiento a través de cada uno de los niveles de la educación haciendo énfasis en las áreas de las matemáticas, ciencias, tecnológicas e ingenierías (STEM).

Para la exploración de los datos y la aplicación de los algoritmos de minería, se hará uso de la herramienta WEKA (entorno para análisis del conocimiento de la Universidad de Waikato) ya que, “proporciona muchos algoritmos diferentes para el análisis de datos y el aprendizaje automático. WEKA es de código abierto y está disponible gratuitamente. También es independiente de la plataforma [13].

Etapa 3: La etapa final del proyecto consiste en la visualización de los indicadores de brecha de género por medio de gráficas y estadísticas en un aplicativo web, de manera que para el usuario final los resultados sean claros y de fácil comprensión. En la construcción de esta herramienta se implementará la metodología de desarrollo ágil SCRUM, ya que es un modelo general de gestión de entornos de producción basados en rutinas, donde el factor más importante es el conocimiento aplicado de las personas que hacen parte del proyecto, aún más primordial que el mismo proceso tecnológico [14]. Teniendo en cuenta que, su importancia radica también en las iteraciones que permiten el desarrollo incremental del proyecto desde su concepción hasta que se dé por terminado.

Todo el proceso descrito anteriormente no tendría sentido sin que exista una forma de que el usuario final visualice los resultados de ambos análisis, por esta razón se desarrollará un aplicativo web que cumpla con la función de mostrar las gráficas y estadísticas de cada una de las pruebas y en cada uno de los períodos evaluados, permitiendo que esta información pueda ser descargada en forma de reporte tipo PDF.

III. CONCLUSIONES

El uso de minería de datos para el proceso de análisis permite tener acceso a información oculta que con los métodos comúnmente usados de estadística no se puede

obtener, además, facilita el procesamiento de grandes volúmenes de datos.

El análisis predictivo se ocupa principalmente de predecir o anticipar resultados futuros basados en minería de datos existentes. Esta información se convierte en importante a la hora de ser analizada porque permite comprender su comportamiento futuro, o en términos coloquiales lo que puede suceder. Esta capacidad de predecir da un poder inmenso de planificación y brinda la posibilidad de solución a diferentes problemáticas ya que entiende cuál es la conducta de una o varias cosas en distintos escenarios de evaluación.

Gracias a las técnicas de minería de datos, en este caso, de agrupamiento y clasificación ad hoc se puede generar modelos descriptivos y predictivos consistentes de la realidad observada.

Con el desarrollo de esta investigación se contribuirá a futuros estudios sobre brecha de género, recomendando así hacer un comparativo de estas pruebas (Saber) con las realizadas a nivel internacional como las PISA.

WEKA facilita la aplicación de los algoritmos de minería de datos siendo una herramienta fiable y de fácil uso para el usuario.

REFERENCIAS

- [1] Semana, "¿Por qué las mujeres siguen ganando menos?", 2016.
- [2] "La Paridad de Género", Unesco, 2019. [Online]. Available: <https://gem-report-2016.unesco.org/es/chapter/la-paridad-de-genero/>. [Accessed: 01- Oct- 2019].
- [3] ONU Mujeres, UNDP, UNFPA, and USAID, De los Objetivos de Desarrollo del Milenio a los Objetivos de Desarrollo Sostenible, no. 88. 2016.
- [4] M. Junguito, "El empoderamiento de las mujeres tiene que empezar desde la infancia", Semana.com, 2019. [Online]. Available: <https://www.semana.com/Item/ArticleAsync/615282?nextId=617220>. [Accessed: 11- Oct- 2019].
- [5] T. Devitt, "Culture, not biology, underpins math gender gap", *University of Wisconsin–Madison*, 2009.
- [6] V. López-Bassols, M. Grazi, C. Guillard, and M. Salazar, "Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe.," *Banco Interam. Desarro.*, p. 52, 2018
- [7] C. Cárcamo and J. Mola, "Diferencias por sexo en el desempeño académico en Colombia: Un análisis regional," *Econ. Región*, vol. 6, no. 1, pp. 133–169, 2012
- [8] C. Pérez López and D. Santín González, *Minería de datos*. [Madrid]: Paraninfo Cengage Learning, 2008.
- [9] D. Abbott, *Applied predictive analytics*. 2014.
- [10] C. Espino Timón, "Análisis predictivo técnicas y modelos utilizados y aplicaciones del mismo," p. 65, 2017.
- [11] R. Timaran and J. Jiménez, "Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de

- Instituciones de Educación Superior con CRISP-DM,”
Form. Univ., pp. 1–19, 2014.
- [12] L. Nel Quezada, *Estadística con SPSS 22*. Perú:
Macro E.I.R.L, 2015.
- [13] S. Singhal and M. Jena, “A Study on WEKA Tool for
Data Preprocessing , Classification and Clustering,” *Int. J.
Innov. Technol. Explor. Eng.*, vol. 2, no. 6, pp. 250–253,
2013.
- [14] J. Palacio and C. Ruata, *Gestión de proyectos
Scrum Manager*. 2011.

ANEXO D Artículo CIETA

INFLUENCE OF GENDER GAPS IN PRIMARY BASIC EDUCATION

INFLUENCIA DE LAS BRECHAS DE GÉNERO EN LA EDUCACIÓN BÁSICA PRIMARIA

Est. Corin J. Viracacha Viracacha, Ing. Cesar Y. Barahona Rodriguez

Universidad de Cundinamarca, Facultad de Ingeniería.
Grupo de Investigación de Sistemas y Tecnología Facativá (GISTFA)
Calle 14 con Avenida 15. Facativá, Cundinamarca, Colombia.
Tel.: (+571) 8920706
E-mail: {cbarahona, cviracacha}@ucundinamarca.edu.co

Abstract: Women are a fundamental entity in the development of scientific research, due to gender stereotypes and historical background, there are global labor, economic and political differences that impede social growth, and this is largely due to models of school education, since sexist biases still persist in these learning processes. For this reason, this research seeks to find aspects related to gender gaps in the country within primary education, for this purpose the fifth knowledge tests will be analyzed, where the socioeconomic information and the score of each thematic nucleus will be related.

Keywords: *Gender gaps, data mining, clustering, education*

Resumen: Las mujeres son un ente primordial en el desarrollo de investigaciones científicas, debido a los estereotipos de género y a los antecedentes históricos, existen a nivel mundial diferencias laborales, económicas y políticas que impiden el crecimiento social, esto en gran parte se debe a los modelos de formación escolar, ya que en estos procesos de aprendizaje aún persisten sesgos sexistas. Por tal motivo, esta investigación busca encontrar aspectos relacionados con las brechas de género en el país dentro de la educación primaria, para ello se analizarán las pruebas saber quinto en donde se relacionará la información socioeconómica y el puntaje de cada núcleo temático.

Palabras clave: Brechas de género, minería de datos, clustering, educación.

1. INTRODUCCION

La educación es clave fundamental en el crecimiento e innovación científica de un país, es por ello que los centros educativos juegan un papel importante en la formación, el proceso de aprendizaje y la construcción de saberes. Por ello, la educación primaria busca desarrollar en los niños competencias lingüísticas, digitales y de pensamiento lógico matemático; para

posteriormente involucrarlas en la formación secundaria y superior.

De acuerdo con el Ministerio de Educación son pocas las mujeres que acceden a programas relacionados con áreas del STEM (Science, Technology, Engineering and Mathematics) probablemente por resultados académicos obtenidos durante la escuela o por factores externos tales como aspectos sociales, familiares y económicos.

Uno de los indicadores que permite conocer la problemática de las brechas de género en el país son las pruebas ICFES, para los investigadores estas pruebas no solo analizan los resultados académicos sino que además permiten relacionar dichos resultados con la información socioeconómica, familiar y personal del estudiante, con el fin, de determinar qué factores influyen en los procesos de aprendizaje.

2. PROCESO DE SELECCIÓN Y ANALISIS DE INDICADORES

2.1. Etapa de selección de datos

A partir de los bancos de datos proporcionados por el ICFES se extrajeron los resultados de las pruebas 5 del año 2017, las cuales cuentan con 54 variables relacionadas con aspectos socioeconómicos, familiares, institucionales y sociodemográficos; que posteriormente serán tratadas para generar una base de datos más consolidada.

Con la información extraída se realiza el proceso de carga al gestor de base de datos PostgreSQL, ya que este permite el manejo y la manipulación de gran volumen de datos; posteriormente a la carga se realizó el proceso de transformación y adecuación (Fig. 1), teniendo en cuenta ciertos parámetros en cuanto al comportamiento de la información.



Fig. 1. Proceso de tratamiento de datos
Fuente: elaboración propia.

2.2. Etapa de procesamiento de los datos

Una vez almacenada la información se realiza el proceso de limpieza, el cual incluye la eliminación de datos nulos y atributos irrelevantes e inconsistentes para el análisis, tales datos como: información del colegio, municipio o departamento; todo este proceso se realiza con el propósito de obtener un almacén de datos más consistente. Adicionalmente se crean índices y se realizan particiones de la tabla, esto para agilizar el

proceso de búsqueda, filtrado y consulta de los datos.

De esta manera se crea una tabla consolidada (Tabla 1), en donde se tuvieron en cuenta de las 54 variables inicialmente obtenidas; una de tipo personal; 4 de tipo socioeconómico y 2 relacionadas con aspectos académicos.

Tabla 1: Tabla consolidada en base de datos

Nombre de la tabla	RESULTADOS_QUINTO
N° de registros	369.315
Variable	Descripción
estu_genero	Genero del estudiante
fami_educacionpadre	Educación del padre
fami_educacionmadre	Educación de la madre
fami_tienecomputador	El estudiante posee computador
fami_tieneconsola videojuegos	El estudiante posee consola de videojuegos
punt_lenguaje	Puntaje en el área de lenguaje
punt_matematicas	Puntaje en el área de matemáticas

Fuente: Elaboración propia

2.3. Etapa de Minería de Datos

Por medio de la minería de datos de la educación se busca descubrir patrones que caractericen el comportamiento de los estudiantes en base a sus logros, evaluaciones y dominios del contenido, con el objetivo de diseñar estrategias educativas por medio de las tecnologías de la información y la comunicación (Hidalgo, 2018).

Dentro de esta etapa se busca por medio de técnicas de Clustering encontrar patrones o grupos interesantes de un conjunto de datos a partir de unos conglomerados (Cluster) con características similares (Carrillo, 2018). Por consiguiente el algoritmo seleccionado como más apropiado para determinar los indicadores es K-means (agrupación), el cual se encargara de clasificar los datos de acuerdo a las características más relevantes. Para la implementación de dicho algoritmo se utilizó WEKA, plataforma de software para aprendizaje automático y minería de datos.

3. RESULTADOS

3.1. Resultados de indicadores socio-económicos

Para Cervini, Dari y Quiroz (2015) el nivel socioeconómico afecta la brecha de género, de manera que los niveles más bajos, tienen un efecto diferencial en el rendimiento entre niños y niñas. Teniendo en cuenta el análisis de cada una de las variables seleccionadas, se muestra a continuación algunos de los indicadores más representativos dentro de la investigación:

3.1.1. Educación de los padres

En cuanto a la educación de los padres se determina que es uno de los factores más influyentes en el rendimiento académico de los niños (Fig. 2).

De manera que la formación académica demuestra que más del 50% de los padres de los niños logran terminar su bachillerato mientras que para las niñas, ellos tan solo alcanzan a terminar la primaria. Lo cual corrobora que los padres juegan un papel fundamental en el desarrollo de aptitudes sirviendo de modelo a sus hijos (Ramirez *et al.*, 2015).

Final cluster centroids:

Attribute	Full Data	Cluster#	
		0	1
	(369315.0)	(265355.0)	(103960.0)
estu_genero	Hombre	Hombre	Mujer
fami_educacionpadre	Bachillerato	Bachillerato	Primaria

Fig. 2. Educación de los padres, con dos clúster.
Fuente: elaboración propia.

Con respecto a la educación de la madre se obtienen porcentajes similares (Fig. 3), concluyendo que gran parte de las madres de los estudiantes logran terminar el bachillerato, en comparación con los resultados de los padres en donde se evidencia que a pesar de algunos estereotipos de la sociedad ellas han demostrado por medio de sus resultados ser modelos de inspiración para sus hijos.

Final cluster centroids:

Attribute	Full Data	Cluster#	
		0	1
	(369315.0)	(188822.0)	(180493.0)
estu_genero	Hombre	Hombre	Mujer
fami_educacionmadre	Bachillerato	Bachillerato	Bachillerato

Fig. 3 Educación de las madres, con dos clúster.
Fuente: elaboración propia.

3.1.2. Acceso a tecnologías

En estos indicadores se logra encontrar información de una posible causa por la cual las niñas obtienen resultados más bajos en algunas áreas y del porque no eligen estudiar carreras relacionadas con las áreas del STEM.

3.1.2.1. Computador

En cuanto al análisis de este indicador se puede evidenciar que en un 74% las niñas no poseen computador mientras que el 26% de los niños lo poseen (Fig. 4). Este es uno de los datos que permiten corroborar el por qué existen brechas de genero digital, de acuerdo a los autores Zapata, Arráiza, Silva y Solares (2015) indican que desde el siglo XX las mujeres tienen menos tiempo de explorar el ciberespacio y de aumentar sus conocimientos tecnológicos, adicionalmente ellos corroboran que aún persiste un paradigma mental en que “los muchachos inventan las cosas y las muchachas usan lo que ellos inventan”, lo que conlleva a un ciberespacio en donde se refleja la sensibilidad masculina al punto de excluir y denigrar lo femenino. Por tal motivo los autores indican que esto conduce al empobrecimiento del contacto con las TIC durante la infancia y la adolescencia así como la generación de un ambiente hostil durante la selección de carreras tecnológicas por parte de las mujeres.

Final cluster centroids:

Attribute	Full Data	Cluster#	
		0	1
	(369315.0)	(245684.0)	(123631.0)
estu_genero	Hombre	Mujer	Hombre
fami_tienecomputador	Sí	No	Sí

Fig. 4. Acceso al computador, con dos clúster.
Fuente: elaboración propia.

3.1.2.2. Consola de Video-juegos

De acuerdo a la clasificación con dos clúster se demuestra que las mujeres en un 74% no poseen consola de video-juegos mientras que en un 26% los hombres poseen este elemento (Fig. 5).

De acuerdo a Rivera y Torres (2018) los video-juegos permiten mejorar el pensamiento lógico y la resolución de problemas, es decir las personas que usan este artefacto tecnológico tienen la capacidad de organizar procesos, crear soluciones a partir de estrategias y generar un pensamiento lógico matemático fuerte. Adicionalmente fortalecen la alfabetización digital y las habilidades de socialización.

```
Final cluster centroids:
Attribute          Full Data      Cluster#
                  (369315.0) (274022.0) (95293.0)
=====
estu_genero       Hombre  Mujer  Hombre
fami_tieneconsolavideojuegos  No      No      Sí
```

Fig. 5. Acceso a consola de video juegos, con dos clúster.

Fuente: elaboración propia.

3.2. Aspectos Académicos

3.2.1. Resultados en Matemáticas

Con respecto a los resultados en matemáticas se evidencia que los niños presentan mejores resultados que las niñas (Fig. 6).

```
Final cluster centroids:
Attribute          Full Data      Cluster#
                  (369315.0) (115390.0) (73289.0) (180636.0)
=====
estu_genero       Hombre  Hombre  Hombre  Mujer
punt_matematicas  291.6883  251.5117  358.031  290.4359
```

Fig. 6. Resultados en matemáticas, con dos clúster.

Fuente: elaboración propia.

Por otro lado, Campos y Santillán (2016) indican que la brecha de género en matemáticas es atribuida a actividades de juego durante la niñez. Por consiguiente los niños tienden a jugar con videojuegos de acción o con juguetes de construcción, lo cual se ha demostrado que este ayuda a mejorar habilidades en matemáticas, otro

aspecto a mencionar es el estereotipo de que las mujeres no tienen buenas habilidades en matemáticas y por tanto relajan el estudio de la materia desde edades tempranas.

3.2.2. Resultados en Lenguaje

En cuanto a las pruebas de lenguaje se observa que las niñas obtienen mejores resultados que los niños (Fig. 7).

```
Final cluster centroids:
Attribute          Full Data      Cluster#
                  (369315.0) (180644.0) (188671.0)
=====
estu_genero       Hombre  Mujer  Hombre
punt_lenguaje     306.3429  312.0636  300.8657
```

Fig. 7. Resultados en lenguaje, con dos clúster.

Fuente: elaboración propia.

De acuerdo a Campos y Santillán (2016) indican que las niñas tienen deficiencias en matemáticas pero se desempeñan mejor en lectura ya que ellas son más propensas a desempeñar actividades relacionadas con la lectura fuera de la clase, otra de las posibles causas son las bajas expectativas que tienen los padres por parte de los hijos en el rendimiento de lectura.

4. CONCLUSIONES

Realizando un comparativo con los resultados de las pruebas PISA(2012) se evidencia que los niños siguen liderando las pruebas de matemáticas con una diferencia de 25 puntos a favor de ellos, mientras que las niñas obtienen mejores resultados en las pruebas de lenguaje con una diferencia de 18 puntos a favor de ellas, de acuerdo a lo resultados obtenidos dentro de la investigación y en comparación a las pruebas PISA se demuestra como el acceso a ciertas tecnologías influyen en el desarrollo de habilidades, capacidades y la creación de conocimientos, adicionalmente de cómo estos aspectos se ven involucrados en la selección de carreras profesionales.

En relación a los puntajes del ICFES se demuestra que existe una desigualdad entre niñas y niños que se amplía en los puntajes de matemáticas, por aspectos no observables como las expectativas de los padres y el ambiente escolar.

A partir de esta investigación se demuestra que los estereotipos surgen dentro de la familia, pues al

considerar que las niñas no son aptas para el uso de artefactos tecnológicos, sino para el uso de juguetes como muñecas o cocinas, dicha estigmatización de la sociedad ha conllevado a ver a la mujer como la encargada del hogar y de los hijos, y no como un ente de gran potencial para la investigación y el desarrollo científico.

SITIOS WEB

Hidalgo, B. (2018). *Minería de datos en los sistemas de gestión de aprendizaje en la educación universitaria*. Recuperado 10 febrero, 2020, de: <https://dialnet.unirioja.es/servlet/articulo?codigo=6681865>,

Carrillo, J. (2018, 8 marzo). *Big Data - Analítica del aprendizaje y minería de datos aplicados en la Universidad*. Recuperado 15 marzo, 2020, de <http://journalprosciences.com/index.php/ps/article/view/56/260>.

Cervini, R., Dari, N., & Quiroz, S. (2015). *Género y rendimiento escolar en América Latina - Los datos del SERCE: matemática y lectura*. Recuperado 1 marzo, 2020, de https://www.researchgate.net/profile/Ruben_Cervini/publication/326679260_Genero_y_rendimiento_escolar_en_America_Latina_Los_datos_del_SERCE_matematica_y_lectura/links/5b5e2225aca272a2d67459b0/Genero-y-rendimiento-escolar-en-America-Latina-Los-datos-del-SERCE-matematica-y-lectura.pdf

Ramírez, A., Ferrando, M., Sainz, A. (2015). *¿Influyen los Estilos Parentales y la Inteligencia Emocional de los Padres en el Desarrollo Emocional de sus Hijos Escolarizados en 2o Ciclo de Educación Infantil?*. Recuperado 29 febrero, 2020, de http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1578-908X2015000100007

Zapata, Cristian Berrío, Arraíza, Paloma Marín, Silva, Ester Ferreira da, & Soares, Elieth das Chagas. (2017). *Desafios da Inclusão Digital: antecedentes, problemáticas e medição da Exclusão Digital de Género*. *Psicología, Conocimiento y Sociedad*, 7(2), 121-151. <https://dx.doi.org/10.26864/pcs.v7.n2.8>. (12 de Febrero de 2020)

Rivera, E., & Torres, V. (2018). *Videojuegos y habilidades del pensamiento*. Recuperado 16

febrero, 2020, de <http://www.scielo.org.mx/pdf/ride/v8n16/2007-7467-ride-8-16-00267.pdf>

Campos, R., & Santillán, A. (2016). *Análisis de diferencias en puntajes en la prueba ENLACE entre niños y niñas en el sistema escolar mexicano*. Recuperado 14 febrero, 2020, de <http://www.scielo.org.mx/pdf/ee/v31n1/0186-7202-ee-31-01-00065.pdf>

Anexo E Artículo CICI 2020

DETERMINACIÓN DE INDICADORES DE BRECHAS DE GÉNERO Y ANÁLISIS DESCRIPTIVO EN LAS PRUEBAS SABER

Alisson Catalina Celeita Gutierrez ¹[0000-0001-9750-2848], Angie Paola Manrique Ravelo ²[0000-0003-0302-406X], Cesar Yesid Barahona Rodriguez ³[0000-0001-7673-7381] y Corin Jazmin Viracacha Viracacha ⁴[0000-0003-4487-938X]

¹ Universidad de Cundinamarca

² Facatativá, Cundinamarca, Colombia

acceleita@ucundinamarca.edu.co

apaolamanrique@ucundinamarca.edu.co

cbarahona@ucundinamarca.edu.co

cviracacha@ucundinamarca.edu.co

Resumen. Las brechas de género son una problemática que perjudica varios aspectos de la sociedad, por esta razón se tiene como objetivo la identificación de aquellos factores que inciden en la educación del país más exactamente en los resultados de las pruebas realizadas por el ICFES, para ello se implementaron técnicas de minería de datos y algoritmos de análisis de agrupamiento (clustering), generando como parte final del proyecto un aplicativo web donde el usuario podrá visualizar de forma gráfica la información resultante para que sea objeto de estudio de futuras investigaciones.

Palabras clave: Brechas de género, minería de datos, clustering, educación, ICFES

INTRODUCCIÓN

Con el paso de los años han existido brechas de género en los ámbitos sociales, económicos y políticos afectando significativamente el desarrollo humano de todos los países en especial a los que conforman a América Latina y el Caribe. Tal es el caso de la gravedad de la problemática en algunos sectores, que por ejemplo en la participación política de las mujeres en Colombia, tan solo el 55.2% ejercen el derecho al sufragio. En cuanto al ámbito económico [1], para los años comprendidos entre el 2014 y 2017 la participación laboral de las mujeres se ha estancado en un 54% dejando al país dos puntos por debajo del promedio de la región [2].

El panorama en el área educativa tampoco es alentador, puesto que, por ejemplo las mujeres indígenas y rurales son quienes menos educación reciben [3]. Así mismo, la participación femenina para el 2018 en las áreas del STEM (Ciencias, tecnología, ingeniería y matemáticas) disminuyó a un 34.9% con respecto al 2001 la cual se ubica en un 38.4%, demostrando que la brecha de género en la educación superior va aumentando [4].

La manera más sencilla para comprender el comportamiento de las brechas de género es guiarse de factores en específico o en áreas puntuales, por esta razón, haciendo uso de los resultados de las pruebas estandarizadas del ICFES se realizó por medio de técnicas de minería de datos un análisis descriptivo donde se evidencie a través de un aplicativo web qué indicadores han incidido en los últimos años de presentación de las pruebas.

1. METODOLOGIA

Para el análisis de las pruebas se hizo uso de la metodología de minería de datos CRISP-DM, la cual tiene varias etapas a seguir. En primera instancia se identificó el problema a tratar, luego se realizó una fase de pre-procesamiento de los datos, la cual inició con la extracción de los resultados del FTP; una vez obtenidos los archivos “.CSV” se procedió a la eliminación de columnas innecesarias y datos nulos, para posteriormente realizar la carga de la información a la base de datos.

Para encontrar los indicadores se hizo uso del algoritmo K-Means, el cual consiste en clasificar los datos y asignarlos a un clúster de acuerdo al centroide más cercano, esto para poder determinar en donde existe mayor concentración y qué relación hay entre las variables seleccionadas para finalmente mostrar dentro del aplicativo una sección con gráficos estadísticos de los resultados hallados.

3. RESULTADOS

Como parte de la información recolectada en la siguiente figura se evidencia la cantidad de datos procesados y analizados por prueba. Es decir, la cantidad de estudiantes por género en cada uno de los niveles evaluados.

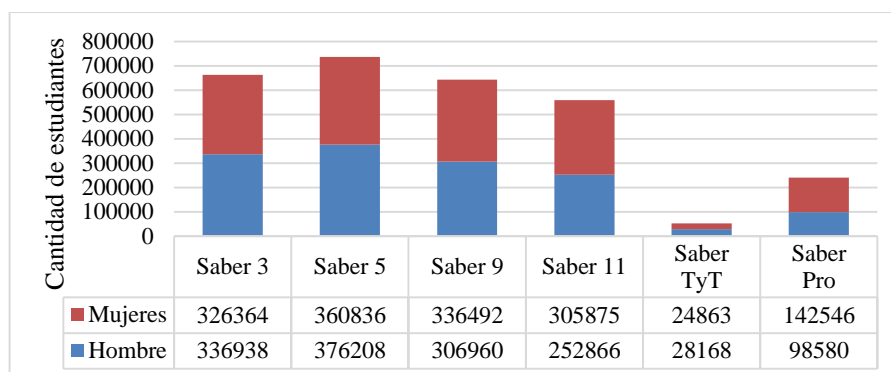


Fig. 1. Cantidad de estudiantes prueba SABER.

De acuerdo a diversos autores para llevar un análisis más organizado se debe categorizar la información en dimensiones académicas, económicas y sociales; por lo anterior, los resultados obtenidos de la etapa descriptiva fueron divididos de esta forma. Para Gálvez los aspectos sociales ejercen una gran influencia en la autodeterminación, seguridad y habilidad para controlar el dinero, el trabajo propio y las relaciones sociales [5]. Mientras que desde el punto de vista académico existe una gran variación en las instituciones educativas, en cuanto a la clasificación profesional y la transmisión académica, que contribuye al aumento de las brechas de género en los países de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) [6].

De todas las variables analizadas y procesadas, finalmente se estableció cuáles son los indicadores con mayor influencia en la brecha de género por prueba.

Tabla 1. Indicadores por dimensiones

Económicos	Académicos	Sociales
Valor pensión	Educación Madre	Estado civil
Estrato de vivienda	Educación Padre	Cabeza de familia
Ingreso familiar	Género del colegio	Número de personas a cargo
Nivel de Sisbén	Jornada del colegio	Área de residencia
Servicio de TV	Dedicación lectura diaria	
Trabaja Actualmente	Área de la carrera	
Tiene Computador	Método de programa	
Tiene Consola de video-juegos		
Tiene internet		
Valor matrícula universidad		
Tipo de pago matrícula.		

- Educación madre y educación padre: Estos indicadores representan los niveles de escolaridad alcanzados por los padres, el cual se clasifica en los siguientes atributos: ninguno, no sabe, no aplica, primaria incompleta, primaria completa, secundaria

incompleta, secundaria completa, técnica y tecnológica incompleta, técnica y tecnológica completa, educación profesional incompleta, educación profesional completa y postgrado.

- Género y jornada del colegio: Estos indicadores definen las características principales del colegio; en primer lugar el género, se clasifica en femenino, masculino y mixto, mientras que la jornada hace referencia al horario de presentación de las clases, clasificado en los siguientes atributos: completa, tarde, mañana, sabatina, noche y única.
- Dedicación lectura diaria: Representa el tiempo de dedicación a la lectura por parte del estudiante, clasificada en rangos de tiempo, de la siguiente manera: 30 minutos o menos, no leo por entretenimiento, entre 30 y 60 minutos, más de 2 horas y entre 1 y 2 horas.
- Área de la carrera: Hace referencia al grupo de clasificación de la carrera que se está cursando. Existen diversas categorías, pero para el análisis se tuvieron en cuenta: Administración y afines, Contaduría y afines, psicología e ingeniería, en el caso de Saber Pro. Y para Saber TyT son: tecnológico en administración y turismo, tecnológico en ingeniería, industria y minas, técnico en administración y turismo, y técnico en ingeniería, industria y minas.
- Método del programa: Hace referencia a la modalidad de estudio del programa, el cual se clasifica en presencial, distancia y virtual.
- Valor pensión: Este indicador representa el rango de valores de la pensión y matrícula, que se paga por el acceso a la educación, está clasificado por rangos de precio de la siguiente manera: entre 87.000 y menos de 120.000, no paga pensión, 250.000 o más, entre 150.000 y menos de 250.000, menos de 87.000, entre 120.000 y menos de 150.000.
- Ingreso familiar: Representa la cantidad de dinero que percibe la familia mensualmente, clasificado en: entre 2 y menos de 3 SMLV, entre 5 y menos de 7 SMLV, menos de 1 SMLV, entre 3 y menos de 5 SMLV, entre 1 y menos de 2 SMLV, entre 7 y menos de 10 SMLV, 10 o más SMLV.
- Nivel del Sisbén y estrato de vivienda: Estos indicadores representan la clasificación de la población de acuerdo a sus condiciones socioeconómicas, para el caso del Sisbén se clasifica en: nivel 1, nivel 2, nivel 3, no está clasificada por el SISBÉN y está clasificada en otro nivel del SISBÉN, en cuanto al estrato sus categorías comprenden desde el estrato 0 hasta el estrato 6.
- Trabaja actualmente: Nos indica si el estudiante se encuentra laborando, clasificada de la siguiente manera: si menos de 20 horas a la semana, si 20 horas o más a la semana y no se encuentra trabajando.

- Servicio de internet y televisión: Nos indica si el estudiante cuenta o no con el servicio de televisión e internet en su hogar.
- Tiene computador y consola video-juegos: Indica si el estudiante cuenta o no con computador y consola de video-juegos en su hogar.
- Valor matricula universidad: Este indicador representa el costo de la matrícula del programa que está cursando el estudiante, clasificado en rangos de pago, de la siguiente manera: entre 5.5 millones y menos de 7 millones, entre 2.5 millones y menos de 4 millones, menos de 500 mil, entre 4 millones y menos de 5.5 millones, entre 500 mil y menos de 1 millón, entre 1 millón y menos de 2.5 millones, Mas de 7 millones y no pago matricula.
- Tipo de pago matricula: Este indicador cuenta con varias opciones de respuesta, pero debido a su afectación nos centramos en el tipo de pago crédito, el cual cuenta con dos opciones de respuesta sí y no.
- Estado civil: Representa la situación de la persona de acuerdo a sus relaciones de pareja, se dividen en 4 opciones: si está casado, soltero, unión libre y separado y/o viudo.
- Cabeza de familia y número de personas a cargo: Determina si el estudiante es o no responsable de la economía de su hogar, en caso de ser responsable se relaciona la cantidad de personas a cargos.
- Área de residencia: Indica la zona de residencia del estudiante, en dicho momento. Sus atributos se clasifican en rural y urbano.

CONCLUSIONES

Gracias al uso de técnicas de minería de datos se pudo descartar variables innecesarias e inconsistentes para el objeto de estudio, determinando así rasgos característicos en común entre los indicadores, y de esta manera clasificarlos en ámbitos académicos, sociales y económicos. A partir de esto se encontró en qué grado afecta cada una de las subcategorías permitiendo establecer una relación entre núcleos temáticos y dichos resultados.

Para futuras investigaciones se recomienda realizar un comparativo entre las pruebas PISA, TIMSS y Saber, el cual permita estimar que otros factores inciden en los resultados académicos, también es recomendable hacer un análisis con las pruebas saber 4°, 6° y 8°, las cuales serán implementadas próximamente por el ICFES.

BIBLIOGRAFIA

1. A. R. L. Francisco, *Colombia: La hora de la paridad política*, vol. 53, no. 9. 2013.
2. ONU Mujeres, UNDP, UNFPA, and USAID, *De los Objetivos de Desarrollo del Milenio a los Objetivos de Desarrollo Sostenible*, no. 88. 2016
3. Lopez, V., Grazi, M., Guillard, C. and Salazar, M., 2018. *Las Brechas De Género En Ciencia, Tecnología E Innovación En América Latina Y El Caribe: Resultados De Una Recolección Piloto Y Propuesta Metodológica Para La Medición*. Banco Interamericano de Desarrollo, accessed 13 June 2020.
4. Patiño, L., 2020. *16 años después, hay más mujeres, pero la brecha no cede*. *El Tiempo*, Available at: <https://www.eltiempo.com/tecnosfera/novedades-tecnologia/brecha-de-genero-mujeres-estudian-ciencia-y-tecnologia-en-colombia-41213>, accessed 9 June 2020.
5. Gálvez, T., 2002. *Aspectos económicos de la Equidad de Género*, *Publicaciones de las Naciones Unidas*, accessed 11 June 2020.
6. Schrøter, J. and Skyt, H., 2013. *Math And Gender: Is Math A Route To A High-Powered Career?*. IZA, accessed 10 June 2020.

ANEXO F Control y seguimiento

- Del 21 Agosto al 11 Septiembre

The screenshot shows the meeting history for 'PROYECTO BRECHAS GENERO'. The interface includes a header with the project name, navigation options (Chat, Archivos, Notas de la reunión), and a 'Unirse' button. The meeting list is as follows:

- 21 de agosto de 2020: CORIN JAZMIN VIRACACHA VIRACACHA y 2 personas más se unieron a la reunión.
- 21/08 12:04 p.m.: Reunión finalizada 1 h 1 min.
- 4 de septiembre de 2020: Reunión iniciada 04/09 11:05 a.m.
- 04/09 11:48 a.m.: Reunión finalizada 42 min 11 s. This entry includes a download error notification: 'Reunión Descarga vencida'.
- 11 de septiembre de 2020: Reunión finalizada 1 min 12 s 11/09 11:05 a.m.

- Del 18 de septiembre al 9 de octubre

The screenshot shows the meeting history for 'PROYECTO BRECHAS GENERO' from September 18 to October 9, 2020. The interface is consistent with the previous screenshot. The meeting list is as follows:

- 18 de septiembre de 2020: Reunión finalizada 1 h 18 min 18/09 12:18 p.m. This entry includes a download error notification: 'Reunión Descarga vencida'.
- 24 de septiembre de 2020: Reunión finalizada 53 s 24/09 3:33 p.m.
- 2 de octubre de 2020: Reunión finalizada 29 min 12 s 02/10 11:28 a.m.
- 9 de octubre de 2020: Reunión finalizada 25 min 4 s 09/10 11:23 a.m. This entry includes a download error notification: 'Reunión Descarga vencida'.

- Del 16 de octubre al 30 de octubre

The screenshot shows a Microsoft Teams chat window for 'PROYECTO BRECHAS GENERO'. The interface includes a top navigation bar with 'Chat', 'Archivos', 'Notas de la reunión', and '1 más'. A 'Unirse' button is visible in the top right. The chat history is organized by date:

- 16 de octubre de 2020:** A meeting ended at 11:35 a.m. with a duration of 37 minutes and 53 seconds. A notification indicates 'Reunión Descarga vencida' (Meeting download expired).
- 23 de octubre de 2020:** A meeting started at 11:01 a.m. A message from CESAR YESID BARAHONA RODRIGUEZ at 11:02 a.m. says 'dame 1 minuto Corin'. A response at 11:02 a.m. says 'Vale, si señor'.
- 30 de octubre de 2020:** A meeting ended at 11:12 a.m. with a duration of 10 minutes and 40 seconds.

- Del 4 de noviembre al 6 de noviembre

The screenshot shows a Microsoft Teams chat window for 'PROYECTO BRECHAS GENERO'. The interface includes a top navigation bar with 'Chat', 'Archivos', 'Notas de la reunión', and '1 más'. A 'Unirse' button is visible in the top right. The chat history is organized by date:

- 4 de noviembre de 2020:** A message at 12:35 p.m. says 'Hola Profe'. A response at 12:35 p.m. says 'hola'. A message at 12:36 p.m. says 'Su merced esta libre?'. A message at 12:36 p.m. says 'en dos minutos'. A meeting started at 12:42 p.m. A message at 12:42 p.m. says 'ya estoy libre'. A meeting ended at 1:15 p.m. with a duration of 32 minutes and 50 seconds.
- 6 de noviembre de 2020:** A meeting ended at 11:24 a.m. with a duration of 22 minutes and 58 seconds.