

**APLICATIVO WEB PARA EL ANÁLISIS, SELECCIÓN Y ADMISIÓN DE
ASPIRANTES AL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA
UNIVERSIDAD DE CUNDINAMARCA EN LA EXTENSIÓN FACATATIVÁ
UTILIZANDO MODELOS PREDICTIVOS DE MINERÍA DE DATOS**

(FASE 2)

DIEGO ALEXANDER GÓMEZ PINZÓN

UNIVERSIDAD DE CUNDINAMARCA

Facultad de Ingeniería

Programa de Ingeniería de Sistemas

Facatativá, noviembre de 2017

**APLICATIVO WEB PARA EL ANÁLISIS, SELECCIÓN Y ADMISIÓN DE
ASPIRANTES AL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA
UNIVERSIDAD DE CUNDINAMARCA EN LA EXTENSIÓN FACATATIVÁ
UTILIZANDO MODELOS PREDICTIVOS DE MINERÍA DE DATOS
(FASE 2)**

DIEGO ALEXANDER GÓMEZ PINZÓN

Director: Ing. Valenzuela Sabogal Gina Maribel, Especialista en Docencia Universitaria,

Magister en Administración y Planificación Educativa

**GRUPO DE INVESTIGACIÓN DE SISTEMAS Y TECNOLOGÍA DE FACATATIVÁ
(GISTFA)**

UNIVERSIDAD DE CUNDINAMARCA

Facultad de Ingeniería

Programa de Ingeniería de Sistemas

Facatativá, noviembre de 2017

Agradecimientos

A mí mismo, por la voluntad de aprendizaje y la perseverancia durante mis estudios y en la realización del presente proyecto.

A mis padres, que me han apoyado y guiado hacia un destino virtuoso.

A todos los docentes que he tenido desde el inicio de mis estudios, por sus valiosos aportes a mi entendimiento.

A las instituciones de educación que me han acogido y hacen posible que las personas se eduquen y amplíen su conocimiento del mundo.

DIEGO ALEXANDER GÓMEZ PINZÓN

Compromiso de autor

Yo, Diego Alexander Gómez Pinzón con célula de identidad No. 1031153109 y con cód. 461213211 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: _____

Resumen

El presente proyecto (fase II) se enfocó en el desarrollo de una plataforma web para el análisis de la deserción estudiantil universitaria en el programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, extensión Facatativá. Para ello se empleó un modelo matemático de minería de datos seleccionado en la fase 1 del proyecto mediante la metodología de proceso KDD. El modelo se utilizó para la tarea de clasificación de perfiles característicos de estudiantes que constan de cuatro dimensiones propuestas por el Ministerio de Educación Nacional de Colombia (institucional, académica, socio-económica e individual).

Se usó la tecnología de desarrollo Python junto con el framework web Flask y una base de datos relacional SQL, el modelo de clasificación depende de la librería Pandas para el tratamiento de datos y Scikit-Learn para el procesamiento matemático. La plataforma web se construyó usando la tecnología RESTful, esta tecnología permite implementar el modelo de datos y los procedimientos programados como recursos de un servicio web, de forma que se facilita e incrementa la escalabilidad del sistema a través de una interfaz uniforme.

El estudio incluyó datos de 649 perfiles de estudiantes universitarios obtenidos mediante encuestas, estos perfiles se usaron para la obtención del modelo clasificador incorporado en la plataforma web. En la plataforma se incluyeron tablas y gráficos generados dinámicamente para analizar el grado de incidencia en la deserción universitaria de cada una de las variables que componen los perfiles característicos de estudiantes tanto para una persona en particular como en grupos de personas.

Palabras clave: Deserción estudiantil, Minería de datos, KDD, Plataforma web, RESTful

Índice general

Informe de investigación.....	17
Fase 2	17
Línea de investigación	17
Estado del arte.....	17
Planteamiento del problema.....	20
Objetivo general.....	21
Objetivos específicos	22
Objetivos fase 2.....	22
Objetivo general (fase 2).....	22
Objetivos específicos (fase 2)	23
Metodología	23
Scrum	24
Impacto del proyecto.....	28
Marcos de referencia.....	29
Marco histórico	29
Antecedentes de estudios de perfiles universitarios.....	30
Marco teórico	32
Deserción estudiantil.....	32

Minería de datos.....	40
Lenguaje de programación python.....	60
Arquitectura orientada a recursos	63
Marco legal	72
Documentación del software.....	74
Plan de proyecto.....	74
Estimación de recursos	74
Determinación de requerimientos (fase 2).....	75
Especificación del sistema	79
Definición del sistema.....	79
Especificación del diseño.....	81
Modelo de entidad relación.....	81
Dimensión Estática	83
Dimensión Dinámica	85
Dimensión Funcional	104
Dimensión organizacional	122
Diseño de los casos de prueba	123
Resultados	145
Transferencia y divulgación del conocimiento.....	149
Conclusiones y recomendaciones	150

Referencias..... 151

Anexos 155

Lista de tablas

Tabla 1 Determinantes de la deserción estudiantil	34
Tabla 2 Tasa de graduación 2015 por área de conocimiento para los niveles TyT agregada y Universitario	38
Tabla 3 Cantidad de desertores para el programa de ingeniería de sistemas UDEC – Facatativá	40
Tabla 4 Algoritmos y tareas de minería de datos.....	55
Tabla 5 Cronograma de actividades.....	74
Tabla 6 Resumen de recursos por rubros	75
Tabla 7 Requerimientos funcionales.....	75
Tabla 8 Definición del sistema	79
Tabla 9 Descripción del Modelo Entidad Relación	82
Tabla 10 Descripción del diagrama de casos de uso.....	84
Tabla 11 Descripción de los diagramas de secuencia.....	104
Tabla 12 Descripción del diagrama de clases	121
Tabla 13 Descripción del diagrama de despliegue	123
Tabla 14 Caso de prueba iniciar sesión en la plataforma web	123
Tabla 15 Caso de prueba registrar un usuario docente administrador en la aplicación	124
Tabla 16 Caso de prueba modificar los datos de acceso de un usuario autenticado en la aplicación	125
Tabla 17 Caso de prueba consultar usuarios registrados en la plataforma web.....	126
Tabla 18 Caso de prueba eliminar una cuenta de usuario de la plataforma web	127
Tabla 19 Caso de prueba recuperar el acceso a la aplicación usando un mensaje de correo electrónico.....	128

Tabla 20 Caso de prueba agregar un aspirante o estudiante	129
Tabla 21 Caso de prueba actualizar los datos de un aspirante o estudiante.....	131
Tabla 22 Caso de prueba visualizar los datos de un aspirante o estudiante.....	133
Tabla 23 Caso de prueba borrar un aspirante o estudiante	133
Tabla 24 Caso de prueba calcular la posibilidad de deserción de una persona	134
Tabla 25 Caso de prueba importar datos de un grupo de personas.....	135
Tabla 26 Caso de prueba buscar aspirantes o estudiantes.....	136
Tabla 27 Caso de prueba borrar un grupo de personas pertenecientes a un resultado de búsqueda	138
Tabla 28 Caso de prueba entrenar el algoritmo de clasificación	138
Tabla 29 Caso de prueba clasificar un grupo de personas por posibilidad de deserción	140
Tabla 30 Caso de prueba calcular la posibilidad de deserción histórica.....	141
Tabla 31 Caso de prueba exportar datos de perfiles de personas a un archivo de hoja de cálculo	143
Tabla 32 Caso de prueba generar informe sobre deserción	143

Lista de gráficos

Figura 1. Tasa de deserción por cohorte 2015, por nivel de formación.....	36
Figura 2. Tasa de deserción cohorte por departamento 2015	37
Figura 3. Porcentaje de deserción Universidad de Cundinamarca	38
Figura 4. Deserción programas de pregrado I P.A. 2008- I P.A. 2016.....	39
Figura 5. Modelo Entidad Relación	82
Figura 6. Diagrama de casos de uso.....	83
Figura 7. Diagrama de secuencia Autenticarse.....	85
Figura 8. Diagrama de secuencia Registrar Usuario.....	86
Figura 9. Diagrama de secuencia Consultar Usuarios	87
Figura 10. Diagrama de secuencia Editar Usuario.....	88
Figura 11. Diagrama de secuencia Eliminar Usuario	89
Figura 12. Diagrama de secuencia Agregar Persona	90
Figura 13. Diagrama de secuencia Mostrar Persona.....	91
Figura 14. Diagrama de secuencia Editar Persona.....	92
Figura 15. Diagrama de secuencia Borrar Persona.....	93
Figura 16. Diagrama de secuencia Recuperar Acceso.....	94
Figura 17. Diagrama de secuencia Importar	95
Figura 18. Diagrama de secuencia Entrenar	96
Figura 19. Diagrama de secuencia Buscar	97
Figura 20. Diagrama de secuencia Borrar Grupo	98
Figura 21. Diagrama de secuencia Exportar	99
Figura 22. Diagrama de secuencia Clasificar Grupo	100

Figura 23. Diagrama de secuencia Deserción Individual	101
Figura 24. Diagrama de secuencia Deserción Histórica	102
Figura 25. Diagrama de secuencia Informe	103
Figura 26. Diagrama de actividad Autenticarse.....	105
Figura 27. Diagrama de actividad Registrar Usuario.....	106
Figura 28. Diagrama de actividad Consultar Usuarios	107
Figura 29. Diagrama de actividad Editar Usuario	107
Figura 30. Diagrama de actividad Eliminar Usuario	108
Figura 31. Diagrama de actividad Agregar Persona	109
Figura 32. Diagrama de actividad Mostrar Persona.....	109
Figura 33. Diagrama de actividad Editar Persona	110
Figura 34. Diagrama de actividad Borrar Persona.....	111
Figura 35. Diagrama de actividad Recuperar Acceso.....	112
Figura 36. Diagrama de actividad Importar	113
Figura 37. Diagrama de actividad Entrenar	114
Figura 38. Diagrama de actividad Buscar.....	115
Figura 39. Diagrama de actividad Borrar Grupo	115
Figura 40. Diagrama de actividad Exportar	116
Figura 41. Diagrama de actividad Clasificar Grupo	117
Figura 42. Diagrama de actividad Deserción Individual	118
Figura 43. Diagrama de actividad Deserción Histórica	119
Figura 44. Diagrama de actividad Informe	120
Figura 45. Diagrama de clases	121

Figura 46. Diagrama de despliegue	122
Figura 47. Resultado de caso de prueba inicio de sesión.....	124
Figura 48. Resultado de caso de prueba registro de usuario.....	125
Figura 49. Resultado de caso de prueba configurar correo.....	126
Figura 50. Resultado de caso de prueba configurar correo.....	126
Figura 51. Resultado de caso de prueba consulta de usuarios	127
Figura 52. Resultado de caso de prueba eliminación de usuarios.....	128
Figura 53. Resultado de caso de prueba enviar instrucciones de recuperación	129
Figura 54. Resultado de caso de prueba restablecimiento de clave	129
Figura 55. Resultado de caso de prueba registro de un aspirante o estudiante	131
Figura 56. Resultado de caso de prueba edición de un aspirante o estudiante	132
Figura 57. Resultado de caso de prueba visualización de un aspirante o estudiante	133
Figura 58. Resultado de caso de prueba eliminación de un aspirante o estudiante	134
Figura 59. Resultado de caso de prueba determinación de la posibilidad de deserción de un aspirante o estudiante	135
Figura 60. Resultado de caso de prueba importar datos de aspirantes o estudiantes desde un archivo de hoja de cálculo.....	136
Figura 61. Resultado de caso de prueba búsqueda de aspirantes o estudiantes	137
Figura 62. Resultado de caso de prueba eliminación de un grupo de aspirantes o estudiantes ..	138
Figura 63. Resultado de caso de prueba entrenamiento del algoritmo de clasificación	140
Figura 64. Resultado de caso de prueba resultados de entrenamiento.....	140
Figura 65. Resultado de caso de prueba clasificación por posibilidad de deserción de un grupo de aspirantes o estudiantes.....	141

Figura 66. Resultado de caso de prueba deserción histórica	142
Figura 67. Resultado de caso de prueba, datos de una hoja de cálculo exportada por la plataforma	143
Figura 68. Resultado de caso de prueba, informe de deserción generado por la plataforma web MIDUDEC	144
Figura 69. Puntaje accuracy en la plataforma web MIDUDEC.....	146
Figura 70. Curva de aprendizaje extraída de la plataforma web MIDUDEC	147
Figura 71. Matriz de confusión extraída de la plataforma web MIDUDEC.....	147
Figura 72. Árbol de decisión extraído de la plataforma web MIDUDEC	148

Introducción

La deserción universitaria es un problema que presenta la educación superior en Colombia, dicho problema influye obviamente en la tasa de graduación, reduciendo la cantidad de profesionales disponibles en el país, de modo que se ve afectado el desarrollo económico y social.

La presente investigación explora una propuesta para el estudio de la deserción desde el área de Ingeniería de Sistemas haciendo uso de técnicas de minería de datos y siguiendo el procedimiento de descubrimiento de conocimiento en bases de datos (KDD), de esta metodología de proceso se obtiene un modelo ajustado al tipo de problema con un perfil característico, partiendo de los resultados del proceso KDD, se prosigue con el desarrollo de una plataforma web que incorpore el modelo.

El algoritmo resultado del proceso KDD corresponde a la tarea de clasificación, se trata de un Regresor Logístico implementado como recurso en la plataforma web mediante una interfaz RESTful que respeta fielmente el propósito de los métodos del protocolo HTTP y posibilita un despliegue distribuido del sistema.

En el proceso de minería de datos son consideradas variables que pertenecen a cuatro categorías de factores determinantes de la deserción estudiados por el Ministerio de Educación Nacional de Colombia, con el desarrollo de una plataforma web para el análisis de la situación se puede determinar en qué medida incide cada factor determinante de la deserción en el caso particular del programa de ingeniería de sistemas de la Universidad de Cundinamarca en la extensión Facatativá y de ese modo apoyar la toma de decisiones sobre el problema.

Para la construcción de la plataforma web se sigue la metodología de desarrollo ágil de software Scrum y se hace uso de los lenguajes Python, HTML, CSS y Javascript con las

herramientas Bootstrap, JQuery, AJAX, Pandas, Scikit-Learn y Flask. Estas tecnologías en conjunto permiten el desarrollo de potentes sitios web dinámicos que soportan procesos de minería de datos.

Informe de investigación

Fase 2

Plataforma Web para la Recopilación de Datos de Perfiles de Estudiantes y el Análisis de la Deserción Universitaria en el Programa de Ingeniería de Sistemas de la Universidad de Cundinamarca en la extensión Facatativá usando modelos predictivos de minería de datos.

Plataforma web MIDUDEC (Minería de Datos de la Deserción en la Universidad de Cundinamarca)

Línea de investigación

El presente proyecto “Aplicativo Web para el Análisis, Selección y Admisión de Aspirantes al Programa de Ingeniería de Sistemas de la Universidad de Cundinamarca en la Extensión Facatativá Utilizando Modelos Predictivos de Minería de Datos (Fase 2)”, se realiza bajo la línea de investigación de Software, Sistemas emergentes y Nuevas tecnologías definida por la Universidad de Cundinamarca bajo el aval del Comité para el Desarrollo de la Investigación.

Estado del arte

En la actualidad, los seres humanos están habituados a usar tecnologías modernas. El desarrollo de sitios web dinámicos hace más fácil la vida cotidiana de la gente. Para la construcción de sitios web se utilizan ahora una serie de herramientas front-end para el diseño web y herramientas back-end para la lógica subyacente. El back-end funciona generalmente en aplicaciones de servidor, realizadas especialmente en los lenguajes de programación Java, Ruby, Python y PHP que proporcionan información a una base de datos SQL. En el lado del diseño se

dispone de HTML5, CSS3, JavaScript especialmente con la librería JQuery, AJAX y Bootstrap. Python es un lenguaje muy popular para la web y se considera como un lenguaje fuerte de scripting del lado del servidor (Sah, 2013).

Junto con las tecnologías mencionadas, existen arquitecturas de software idóneas para construir una plataforma web moderna. Según Reid, Montgomery, & Drake, (2016), una manera de mejorar la accesibilidad e interoperabilidad de paquetes de software de aprendizaje automático (Machine Learning) es proveerles una interfaz abstracta por medio de servicios web. Una arquitectura orientada a recursos (ROA, por sus siglas en inglés) con transferencia de estado representacional (REST, por sus siglas en inglés) son mecanismos bastante adecuados para esta tarea. Recientemente ha habido un rápido crecimiento en el área de servicios web con Machine Learning, tales como Google Prediction API, OpenTox, BigML, Microsoft Azure ML, Wise.io, entre otros. En estos servicios es consistente el uso de HTTP y JSON en sus APIs REST.

La minería de datos enfocada a la deserción es un tema sobre el cual se han realizado varias investigaciones. En Estados Unidos se desarrolló un modelo predictivo de minería de datos para predecir la posibilidad de regreso a clases de cada alumno que se encontraba matriculado en Silicon Valley (Luan, 2002), en el mismo país se realizó una investigación para utilizar técnicas de minería de datos con el fin de hallar variables y factores que afectan a la deserción de los alumnos (Veitch, 2010). En México, se desarrolló una herramienta para predecir la probabilidad de deserción de los alumnos mientras cursan sus estudios (Orea, 2009). En Argentina, usando inducción en árboles de decisión, se hizo un análisis con el fin de encontrar reglas de comportamiento que contengan variables de incidencia en la retención de estudiantes universitarios para planificar políticas de retención preventiva (Kuna Horacio, García Martínez Ramón, 2010), también en Argentina, se

implementó en MS SQL Server, SPSS y Weka un clasificador de rendimiento académico para detectar patrones determinantes de la deserción estudiantil (Sposito & Etcheverry, 2010).

Los trabajos realizados en minería de datos para tratar el problema de la deserción no sólo están orientados a la presencia de la situación en la educación superior. En España, Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto (2012) realizaron un sistema informático para la predicción del fracaso escolar mediante técnicas de minería de datos, el sistema utilizó técnicas de minería de datos para detectar cuáles son los factores que más influyen para que los estudiantes de enseñanza media o secundaria suspendan o abandonen sus estudios.

Más recientemente, en Venezuela, se realizó un trabajo de investigación denominado “Minería de datos aplicada a la deserción estudiantil. Caso: Licenciatura en Computación de la Universidad del Zulia-NPF” (Marcano & Rodríguez, 2014).

El objetivo de la investigación fue obtener patrones sobre los estudiantes que no han podido concluir sus estudios universitarios, en la Licenciatura en Computación de LUZ- NPF; aplicando para ello la Minería de datos. Corresponde a una investigación de tipo descriptiva de campo y se desarrolló bajo la metodología computacional Crisp-DM, con apoyo de Weka. (Marcano & Rodríguez, 2014, p. 32)

En Colombia, se realizó un proyecto de investigación en la Universidad de La Sabana cuyo objetivo era encontrar por medio de técnicas de minería de datos, en una base de datos de estudiantes, los atributos de mayor incidencia en la deserción dentro de dicha institución (Morales, 2012). También, en la Universidad de Nariño se realizó una investigación cuyo objetivo fue determinar perfiles de bajo rendimiento académico y deserción estudiantil aplicando técnicas de descubrimiento de conocimiento (Pereira, 2009). Así mismo, en la Universidad Simón Bolívar, se

llevó a cabo un análisis de la deserción estudiantil utilizando la herramienta de minería de datos Weka (Azoumana, 2013). Otra investigación realizada en Colombia es la de Amaya, Barrientos, & Santander (2014), se trataba de un modelo predictivo de deserción estudiantil aplicando técnicas de minería de datos; el trabajo buscaba predecir la probabilidad de que un estudiante abandone sus estudios, dadas sus características socio-económicas y académicas. Más recientemente, en la Universidad El Bosque, se hace una investigación para construir un modelo predictivo del rendimiento académico usando arboles de decisión, redes neuronales artificiales y otras técnicas de clasificación (Merchán & Duarte, 2016).

La fase previa del presente proyecto, realizada en la Universidad de Cundinamarca tuvo como objetivo realizar un aplicativo para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas en la extensión de Facatativá, utilizando modelos predictivos de minería de datos; con esa investigación se logró determinar que el método de regresión logística es el mejor entre varios algoritmos de minería de datos evaluados para el problema planteado, también se establecieron los arboles de decisión como método de explicación de las reglas que intervienen en el proceso (Pertuz Arroyave & Chaves Sánchez, 2016).

Todas las investigaciones previas mencionadas contribuyen a una conceptualización de la problemática y de la viabilidad del presente proyecto.

Planteamiento del problema

En las estadísticas de deserción y graduación en la educación superior del ministerio de educación se puede evidenciar que todas las áreas de conocimiento y todos los niveles de formación presentan altos porcentajes de deserción, aunque en los últimos años Colombia ha avanzado en la disminución de la deserción estudiantil todavía sigue siendo un fenómeno de

seguimiento por parte del ministerio de educación nacional. Es de resaltar que el área de ingeniería es una de las que menor tasa de graduación presenta en el artículo (M. de E. de Colombia, 2015).

El abandono de los estudios no es una excepción en la Universidad de Cundinamarca, tampoco lo es en el área de ingeniería, debido a esta situación, se propone desarrollar una plataforma web que clasifique de forma confiable los aspirantes al programa de ingeniería de sistemas en la extensión de Facatativá y seleccione los estudiantes que ingresarán con la menor probabilidad de abandono de los estudios, de ese modo se busca contribuir con la disminución de la deserción estudiantil en la educación superior, que repercutirá en las estadísticas ofrecidas por el Sistema para la Prevención de la Deserción de la Educación Superior (SPADIES) para que el ministerio de educación pueda alcanzar las metas que proponga el gobierno nacional en esta materia, además la Universidad de Cundinamarca podrá ofrecer una mayor tasa de graduación disminuyendo por supuesto los gastos infructuosos de recursos asociados a la deserción estudiantil.

Teniendo en cuenta la relevancia de la realización del proyecto surge la siguiente pregunta: ¿Es posible que una plataforma web que use técnicas de minería de datos clasifique perfiles de aspirantes y estudiantes en cuanto a la deserción en el programa de Ingeniería de Sistemas ofrecido por la Universidad de Cundinamarca en la extensión Facatativá?

Objetivo general

Desarrollar un aplicativo web para el análisis y evaluación de patrones de conocimiento sobre caracterización de perfiles de aspirantes al programa de Ingeniería de Sistemas apoyando las políticas de selección y admisión de la universidad de Cundinamarca extensión Facatativá.

Objetivos específicos

- Diseñar un estándar que permita un compendio de datos referido a variables homogéneas y correlacionadas de perfiles de aspirantes a pregrado.
- Establecer el algoritmo de minería, la tarea y el método apropiados, que permitan obtener el tipo de modelo predictivo, que se acople a los requerimientos del problema en estudio.
- Construir un módulo que permita la recopilación de datos tanto de estudiantes activos, como de aspirantes a ingresar a la carrera, que permita a su vez la alimentación automática de la base de datos del sistema.
- Crear una sección del aplicativo que permita la selección de aspirantes en el periodo académico en el que se desea ejecutar el proceso de admisión, posteriormente a la aplicación del algoritmo establecido.
- Definir un módulo de consulta y actualización de la base de datos en cuanto al estatus de los aspirantes evaluados.
- Describir por medio de gráficos y reportes en formato PDF, tanto el proceso como los resultados del análisis de perfiles para la evaluación e interpretación del experto.

Objetivos fase 2

Objetivo general (fase 2)

Desarrollar una plataforma web para la recopilación de datos, evaluación de patrones de conocimiento y selección de perfiles de aspirantes al programa de ingeniería de sistemas apoyando las políticas de admisión de la universidad de Cundinamarca extensión Facatativá.

Objetivos específicos (fase 2)

- Realizar el modelamiento de la plataforma web.
- Realizar un diseño estático del prototipo de la interfaz de usuario web.
- Construir un servicio que siga los principios REST para la aplicación.
- Desarrollar el consumo de los servicios REST por parte de la interfaz de usuario web.
- Elaborar la documentación correspondiente a la plataforma web.
- Realizar pruebas funcionales para verificar el funcionamiento adecuado de la aplicación.

Metodología

Se propone un proyecto de investigación aplicada, de tipo correlacional con un enfoque cuantitativo. Se trata de una investigación aplicada porque el presente proyecto tendrá como resultado la implementación de teoría propia de técnicas de minería de datos en un producto software para uso en un entorno real, es una investigación de tipo correlacional puesto que se utiliza el grado de relación entre variables independientes, se empleará un enfoque cuantitativo para evaluar la probabilidad de deserción de un aspirante, no obstante, se utilizarán variables cualitativas que serán transformadas para ser analizadas de forma cuantitativa, con lo cual se facilita el tratamiento computacional del problema.

Para la realización de las diferentes actividades que enmarcan este proyecto se pretende utilizar el lenguaje de programación Python porque dispone de librerías que se pueden utilizar para el proceso matemático requerido, también cuenta con frameworks de desarrollo web propicios para la implementación de buenas prácticas de codificación y que ahorran tiempo en la realización de tareas comunes en cualquier plataforma web al estar pre configuradas para ser generadas de manera automática utilizando por lo general un intérprete de comandos. Dentro de los frameworks

disponibles en el lenguaje de programación Python se elige Flask debido a la mayor compatibilidad que presenta con las librerías usadas en la primera fase de este proyecto.

La plataforma hará uso de varias tecnologías para la construcción de sitios web, por el lado del modelo de datos se manejará una base de datos relacional SQL, para el proceso de minería de datos se utilizara el código proporcionado por la primera fase de este proyecto, para comunicar el modelo de datos con la lógica del proyecto, con el usuario y con otros sistemas de información; se propone entonces construir un servicio web con arquitectura orientada a recursos y transferencia de estado representacional debido a las numerosas ventajas que presenta en cuanto a escalabilidad e integración con otros sistemas. La interfaz de usuario se desarrollará usando Bootstrap, un framework para el diseño de sitios y aplicaciones web.

Se usará SCRUM como metodología de desarrollo de software para la construcción de la plataforma; SCRUM hace posible un desarrollo ágil e incremental de la aplicación, lo cual permite la detección y corrección de errores funcionales en la aplicación desde las primeras etapas del desarrollo.

Scrum

Scrum es un marco de trabajo dentro del cual las personas pueden abordar complejos problemas adaptativos, mientras que de manera productiva y creativa se entregan productos del mayor valor posible. Scrum fue desarrollado por Jeff Sutherland en 1993. El marco de Scrum consta de equipos de personas y sus roles, eventos, artefactos y reglas asociadas. Cada componente del marco cumple un propósito específico y es esencial para el éxito y el uso de Scrum (Schwaber & Sutherland, 2013).

Equipo Scrum

Un equipo Scrum está compuesto por (Ribeiro Lima, Castro Freire, & Costa, 2012):

- Equipo de desarrollo (Team): está compuesto por hasta diez desarrolladores donde cada miembro tiene una habilidad específica. Sin embargo, los miembros no tienen restricciones en realizar tareas diferentes a las de su cargo. Así, el equipo se integra más y los miembros de los equipos conocen mejor el software, minimizando el impacto del despido de un miembro.
- Dueño del producto (Product owner): es responsable de la especificación de la funcionalidad del software y resolver cualquier duda que pueda surgir durante el desarrollo. Es el representante del cliente, debe observar el proyecto de cerca y ayudar en la construcción de un software que responda completamente a las necesidades del cliente.
- Maestro de Scrum (Scrum master): es el responsable de dirigir el equipo y evitar los obstáculos que puedan surgir durante el proceso. Un obstáculo es algo que podría impedir que un miembro realice su trabajo. Por ejemplo, las solicitudes para realizar actividades no relacionadas con el proyecto, los problemas en el servidor de prueba, las dificultades con la tecnología y los requisitos no planificados podrían ser ejemplos de obstáculos que podrían causar problemas al sprint.

Sprint

El evento principal de Scrum es el Sprint, una caja de tiempo de un mes o menos durante la cual se crea un incremento de producto hecho, utilizable y potencialmente liberable. Un nuevo Sprint comienza inmediatamente después de la conclusión del Sprint anterior. Los Sprints

contienen y consisten en: planificación de Sprint, Scrums diarios, trabajo de desarrollo, revisión de Sprint y retrospectiva de Sprint (Schwaber & Sutherland, 2013).

- **Planificación de Sprint:** El trabajo que se realizará en el Sprint se planifica en la planificación de Sprint. Este plan es creado entre todo el equipo scrum. La planificación de Sprint se realiza en un plazo máximo de ocho horas para un Sprint de un mes. El equipo de desarrollo pronostica y negocia con el Product Owner las tareas extraídas de una pila de producto (Product Backlog) que se pueden realizar en el Sprint y luego el equipo Scrum define una meta de Sprint. Posteriormente se define cómo se realizarán las tareas con el fin de crear un Sprint Backlog que se descompone en unidades de un día o menos para que el equipo de desarrollo se auto-organice y emprenda el trabajo.
- **Scrum diario:** es un evento de 15 minutos para que el equipo de desarrollo sincronice las actividades y cree un plan para las próximas 24 horas. El Scrum diario se lleva a cabo al mismo tiempo y lugar cada día para reducir la complejidad. Sólo los miembros del equipo de desarrollo participan en el Scrum diario, el Scrum master asegura que se realice la reunión y que tenga una duración de 15 minutos.
- **Revisión de Sprint:** es una reunión informal que se lleva a cabo al final del Sprint para inspeccionar el incremento y adaptar el Product Backlog si es necesario. Tiene una duración de cuatro horas para un Sprint de un mes. En esta reunión, el equipo Scrum y las partes interesadas colaboran para obtener comentarios de cosas que se podrían hacer para optimizar el valor. El resultado de la revisión de Sprint es un Product Backlog revisado que define los elementos probables del Product Backlog para el próximo Sprint.
- **Retrospectiva de Sprint:** La retrospectiva de Sprint es una oportunidad para que el equipo Scrum se inspeccione y cree un plan de mejoras que se implementarán durante el próximo

Sprint. La retrospectiva de Sprint se realiza después de la revisión de Sprint y antes de la próxima planificación de Sprint. Se trata de una reunión de tres horas para Sprints de un mes.

Artefactos de Scrum

Los artefactos de Scrum maximizan el intercambio transparente de información clave. A continuación, se definen los artefactos de Scrum (Schwaber & Sutherland, 2013):

- **Product Backlog:** es una lista ordenada de todo lo que puede ser necesario en el producto y es la única fuente de requisitos para cualquier cambio que se realice en el producto. El Product Owner es responsable del Product Backlog, incluyendo su contenido, disponibilidad y ordenamiento. El Product Backlog evoluciona a medida que lo hace el producto y el ambiente donde se implementa para adaptarse a necesidades cambiantes. El Product Backlog enumera todas las características, funciones, requisitos, mejoras y correcciones que constituyen los cambios que se realizarán en el producto en versiones futuras.
- **Sprint Backlog:** es el conjunto de elementos del Product Backlog seleccionados para el Sprint, además de un plan para entregar el incremento de producto y alcanzar la meta del Sprint. El Sprint Backlog puede ser modificado por el equipo de desarrollo durante el progreso del Sprint.

Incremento: es la suma de todos los elementos del Product Backlog completados durante un Sprint y el valor de los incrementos de todos los Sprints anteriores.

Impacto del proyecto

La deserción de estudiantes en los programas de pregrado es un problema que afecta a muchas instituciones de educación superior en Colombia, incluso a la Universidad de Cundinamarca; este problema repercute en el desarrollo del país debido a la escasez existente de profesionales especialmente en las diferentes áreas de la ingeniería. Algunas universidades cuentan con programas de apoyo a los estudiantes para fomentar su permanencia, pero en varias ocasiones no son suficientes y los alumnos optan por abandonar sus carreras profesionales por diversos motivos atribuidos a inconvenientes externos o internos de las instituciones de educación superior, problemas propios del estudiante y/o a causas personales o familiares (Pertuz Arroyave & Chaves Sánchez, 2016).

Una plataforma web que sirva de apoyo al proceso de selección de aspirantes al programa de ingeniería de sistemas, que se fundamente en modelos de predicción aportados por la técnica de minería de datos y que clasifique con un porcentaje alto de credibilidad los aspirantes de acuerdo a la probabilidad de deserción analizando los datos de sus perfiles, permitirá analizar los índices de deserción como una medida preventiva que puede ahorrar esfuerzos tanto para el estudiante desertor como para la universidad.

Marcos de referencia

Marco histórico

Aproximadamente desde los años sesenta los estadísticos manejaban términos como data fishing, data mining o data archeology con la idea de encontrar relaciones existentes en los datos en bases de datos con ruido. A principios de los años ochenta se empezaron a consolidar los términos de la minería de datos, época donde sólo existían unas pocas empresas dedicadas a este estudio, para el 2002 ya existían más de 100 empresas en todo el mundo que utilizan la minería de datos y las listas de discusión sobre este tema las formaban investigadores de más de ochenta países. Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios (Martinez, 2004).

En los inicios del año 1996, el modelo KDD (Knowledge Discovery in Databases) constituyó el primer modelo aceptado en la comunidad científica que estableció las etapas principales de un proyecto de explotación de información. En su versión completa, KDD está formado por nueve etapas, donde la primera es el entendimiento del negocio. Formalmente el modelo establece que la minería de datos es la etapa dentro del proceso en la cual se realiza la extracción de patrones a partir de los datos. Sin embargo, actualmente, en la comunidad científica y en la literatura, el término KDD y minería de datos se utilizan indistintamente para hacer referencia al proceso completo de descubrimiento de conocimiento. A partir del año 2000, con el gran crecimiento en el área de la minería de datos, surgen tres nuevos modelos que plantean un enfoque sistemático para llevar a cabo el proceso: SEMMA, CRISP-DM y Catalyst (también conocido como P3TQ) (Moine, Gordillo, & Haedo, 2011).

Antecedentes de estudios de perfiles universitarios

Es relevante para las instituciones de educación superior tener el mayor conocimiento posible del estudiante universitario para poder tomar decisiones sobre la apertura de un programa académico o establecer políticas de bienestar para los estudiantes en un programa existente, por tal motivo, la Asociación Colombiana de Universidades (ASCUN) emprendió a finales de la década de los años 80 un programa para definir las características del universitario colombiano y organizar un sistema de información sobre su perfil (Pinzon Cadena, 2011).

En el año 2000 se publicó una investigación titulada: “Caracterización del estudiante universitario de Santa fe de Bogotá”, que fue desarrollada con la colaboración de 8 instituciones adscritas a ASCUN. Posteriormente, otras universidades han hecho estudios para caracterizar los estudiantes de su institución en general y/o caracterizar los estudiantes de un programa académico ofrecido. Todas las investigaciones se suelen realizar mediante estudios analíticos descriptivos, utilizando como instrumentos encuestas, tomando muestras representativas de la totalidad de estudiantes y haciendo uso de análisis estadísticos y relacionales (Pinzon Cadena, 2011).

Con la Ley 749 del año 2002 el Ministerio de Educación Nacional de Colombia propuso disminuir la deserción estudiantil en la educación superior como parte de la estrategia planteada para aumentar la cobertura, la calidad y la eficiencia educativas. Una de las acciones tomadas fue poner en marcha el Sistema para la Prevención de la Deserción en la Educación Superior (SPADIES) ante la necesidad de contar con una visión sectorial e integrada de la problemática de la deserción, a partir de la cual se pudiera disponer de una conceptualización, una medición y una metodología de seguimiento del fenómeno aplicables a todas las Instituciones de Educación Superior del país (Emanuel Chengu, 2007).

En 2004 se hizo un concurso público de méritos que fue otorgado al Centro de Estudios sobre Desarrollo Económico (CEDE) de la Universidad de los Andes. Este último desarrolló la herramienta SPADIES, la implementó desde noviembre de 2004 hasta noviembre de 2011, y la instaló en las instituciones de educación superior. A lo largo de los años, a la herramienta se le han hecho ajustes con el fin de mejorar la calidad y la utilidad de la información obtenida. Si bien la razón fundamental para la existencia del SPADIES es evitar la deserción estudiantil, también se ha visto la necesidad de identificar casos de rezago estudiantil, es decir, estudiantes que no desertan, sino que permanecen matriculados sin graduarse (Emanuel Chengu, 2007).

En las universidades colombianas se han realizado diversos estudios sobre el perfil de los estudiantes de educación superior, uno de ellos es el estudio denominado “voces y rastros de jóvenes javerianos” de la Pontificia Universidad Javeriana, estudio de tipo exploratorio que abarca el perfil socio-demográfico y cultural, información académico y cognoscitivo, socio afectivo y sexual, ético-moral y religioso, participativo y gremial, político e ideológico, salud-tiempo libre y deportivo; la Universidad EAFIT realizó el estudio “Perfil Socioeconómico del Estudiantado de EAFIT”, investigación que tuvo en cuenta características demográficas, aspectos socioeconómicos y oportunidades sociales, indicadores de comportamiento y comportamiento académico. En la Universidad Autónoma de Cali en el año 2007, la Oficina de Planeación y desarrollo Institucional Bienestar Universitario realizó una investigación sobre el perfil socioeconómico del estudiante y se tomó las variables demográficas, biográficas, socioeconómicas, académicas y uso del tiempo libre. Por último, la Universidad Javeriana, publicó un estudio llamado “Factores individuales que afectan la Demanda de Educación Superior de Ingenierías: Caso de la Pontificia Universidad de Cali”, donde se estudian las razones fundamentales en la elección de la carrera (Pinzon Cadena, 2011).

Marco teórico

Deserción estudiantil

La educación superior enfrenta cambios asociados a su misión educativa en la sociedad. Estos cambios se expresan en una situación muy particular, la deserción estudiantil. En Colombia, el acceso a la educación superior en las últimas décadas tiene como tendencia la universalización, es decir, una masificación de políticas económicas direccionadas por el Estado que tendrían como función financiar los estudios universitarios de diferentes capas poblacionales. Sin embargo, esto contrasta con el hecho evidente y además problemático de altas tasas de estudiantes que no logran finalizar satisfactoriamente su formación académica. La deserción estudiantil es un síntoma de una problemática educativa a nivel estructural, relacionado con las condiciones sociales en las que se genera el discurso educativo, no sólo de parte de las instituciones de educación superior, sino también, de los estudiantes (Bravo Castillo & Mejía Giraldo, 2010).

Se puede entender la deserción como una situación a la que se enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo, considerándose como desertor a aquel individuo que siendo estudiante de una institución de educación superior no presenta actividad académica durante dos semestres académicos consecutivos, lo cual equivale a un año de inactividad académica. Esta es la definición que ha adoptado el Ministerio de Educación Nacional de Colombia para la medición y seguimiento de la problemática (Guzmán Ruiz, Muriel Durán, & Franco Gallego, 2009).

A partir de tal definición se pueden diferenciar dos tipos de abandono en estudiantes universitarios: uno con respecto al tiempo y otro con respecto al espacio (Guzmán Ruiz et al., 2009).

La deserción con respecto al tiempo se clasifica a su vez en (Guzmán Ruiz et al., 2009):

- Deserción precoz: individuo que habiendo sido admitido por la institución de educación superior no se matricula.
- Deserción temprana: individuo que abandona sus estudios en los primeros semestres del programa.
- Deserción tardía: individuo que abandona los estudios en los últimos semestres.

Por otro lado, la deserción con respecto al espacio se clasifica en (Guzmán Ruiz et al., 2009):

- Deserción interna o del programa académico: Ocurre cuando el estudiante decide cambiar de programa académico por otro que ofrece la misma institución universitaria.
- Deserción institucional: Ocurre cuando el estudiante cambia de institución universitaria.
- Deserción del sistema educativo: Ocurre cuando el estudiante definitivamente se retira del programa académico y de toda actividad estudiantil.

Factores determinantes

Basados en la literatura sobre el tema, puede decirse que hoy en día existe consenso al afirmar que la deserción estudiantil es el resultado del efecto no de una sola categoría, sino del efecto individual y de la interacción de diferentes categorías de factores, a saber: factores individuales, factores académicos, factores institucionales y factores socioeconómicos. A partir de los diferentes estudios previos, las principales variables consideradas dentro de cada una de las categorías se resumen a continuación en la Tabla 1 (Universidad del los Andes, 2002).

Tabla 1
Determinantes de la deserción estudiantil

Individuales	Socioeconómicos	Académicos	Institucionales
- Edad, género, y estado civil	- Estrato social	- Orientación profesional	- Normatividad académica
- Calamidad y/o problema doméstico	- Situación laboral del estudiante	- Tipo de colegio de secundaria	- Becas y formas de financiamiento
- Integración social	- Situación laboral de los padres	- Rendimiento académico superior	- Recursos universitarios
- Expectativas no satisfechas	- Dependencia económica	- Métodos de estudio	- Relaciones con el profesorado y con demás estudiantes
- Incompatibilidad horaria con actividades extra académicas	- Personas a cargo	- Calificación en el examen de admisión	- Grado de compromiso con la institución educativa
	- Nivel educativo de los padres	- Insatisfacción con el programa académico	- Calidad del programa
	- Entorno familiar	- Carga académica (número de materias al semestre)	
	- Entorno macroeconómico del país	- Repitencia	

Fuente: (Universidad del los Andes, 2002)

Deserción estudiantil en Colombia

En el año 2003 comienzan a desarrollarse en Colombia investigaciones que, partiendo de la revisión exhaustiva de la literatura existente, construyeran el estado del arte sobre la deserción estudiantil e identificaran los cuatro grupos de factores y las variables que dentro de cada grupo permiten explicar empíricamente sus causas. Esta aproximación teórica permitió establecer diferencias en cuanto al estudio de la deserción dependiendo del tiempo y del espacio (Guzmán Ruiz et al., 2009).

Se pueden destacar tres trabajos claves en el entendimiento del problema de la deserción estudiantil en el ámbito nacional. El primero, realizado por la Universidad Nacional de Colombia y el Instituto Colombiano para el Fomento de la Educación Superior (ICFES), el cual contribuyó a la comprensión teórica y conceptual del fenómeno; el segundo, hecho por la Universidad de Antioquia, en el que se implementaron técnicas estadísticas adecuadas para estudiar la deserción como un problema dinámico y, el tercero, desarrollado por el Ministerio de Educación Nacional

con apoyo de la Universidad de los Andes, instituciones que además de hacer un análisis del riesgo de deserción, desarrollaron el Sistema de Prevención de la Deserción en Educación Superior o SPADIES, que permite a cada institución identificar y clasificar a los estudiantes en riesgo de deserción, dado un grupo determinado de variables. Este estudio se basó en los aportes de los dos estudios antes mencionados (Guzmán Ruiz et al., 2009).

Estudios recientes sobre la deserción en la educación superior en Colombia producen las siguientes conclusiones (Fundación Universitaria Luis Amigó, 2013):

- El principal factor determinante del abandono de estudios se sitúa en la dimensión académica: está asociado al potencial cultural y académico con el cual ingresan los estudiantes a la educación superior.
- Los factores financieros y socioeconómicos están a continuación, seguidos por los institucionales y los de orientación vocacional y profesional.
- Gran parte de los alumnos abandona sus estudios en los primeros semestres (alrededor de un 60% del total de desertores), especialmente por causas académicas y de orientación profesional y vocacional.
- Están ingresando más estudiantes a la educación superior, pero en condiciones que los exponen a mayores retos en el campo académico, económico y social.
- Se observa una relación directa entre mayores ingresos y mejores resultados académicos. Son determinantes para obtener buenos resultados académicos: el nivel de ingresos familiares, poseer vivienda propia y tener un trabajo.
- Cerca de la mitad de la población de estudiantes que ingresa a la Universidad proviene de familias con menores antecedentes académicos.

- Existe un porcentaje creciente de estudiantes maduros que enfrentan compromisos personales, económicos y laborales distintos dentro de su núcleo familiar.

La metodología de seguimiento y medición de la deserción adoptada por el Ministerio de Educación Nacional es usada en el Sistema de Prevención de la Deserción en Educación Superior (Fundación Universitaria Luis Amigó, 2013). El SPADIES permite medir la deserción a través de dos metodologías, una es la deserción por periodo, la cual hace parte de los indicadores del plan de desarrollo, que corresponde a la proporción de estudiantes que estando matriculados dos semestres atrás son clasificados como desertores un año después. La otra mide la deserción por cohorte que contabiliza la deserción acumulada en cada semestre para un grupo de estudiantes que ingresaron a primer curso en un mismo periodo académico (cohorte), la cual observa para el nivel universitario al décimo semestre (M. de E. de Colombia, 2015).

El SPADIES monitorea la deserción en los diferentes niveles de formación de la educación superior en Colombia, en la Figura 1 se observa la tasa de deserción por cohorte en Colombia por nivel de formación, la gráfica incluye la tasa de deserción para el nivel de formación universitaria en todo el país, con un porcentaje de desertores de 46,1 %.

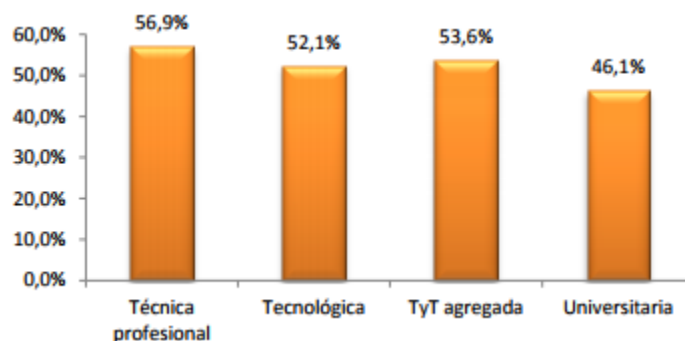


Figura 1. Tasa de deserción por cohorte 2015, por nivel de formación

Fuente: SPADIES corte abril 2016. En: (M. de E. de Colombia, 2015)

La Figura 2 muestra la tasa de deserción por departamento, en este caso, Cundinamarca presenta un porcentaje de deserción de 45,1 %, valor inferior a la media nacional, pero a la vez preocupante, de hecho, se puede afirmar que cerca de la mitad de los estudiantes que ingresan a los programas de educación superior se retiran sin culminar el proceso, tanto a nivel nacional como en el departamento de Cundinamarca.

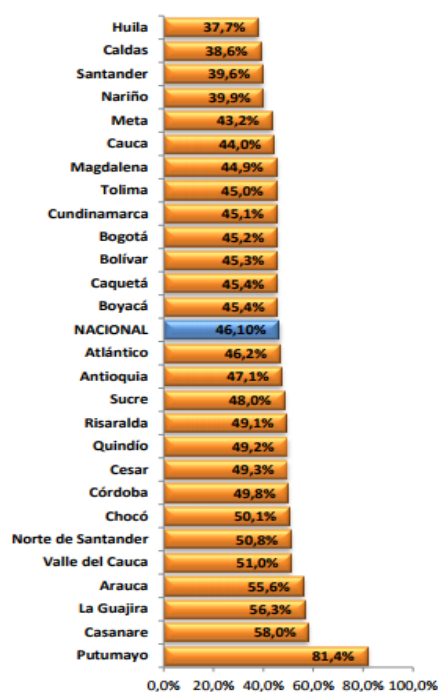


Figura 2. Tasa de deserción cohorte por departamento 2015

Fuente: SPADIES corte abril 2016. En: (M. de E. de Colombia, 2015)

Con la información reportada al SPADIES de las instituciones de educación superior, igualmente se analiza la tasa de graduación, la cual contabiliza el número total de graduados, sobre el total acumulado de los estudiantes que ingresaron a primer curso (M. de E. de Colombia, 2015). En la Tabla 2 se observa la tasa de graduación según el área de conocimiento en Colombia, en la tabla, Ingeniería de Sistemas se encuentra en el área de Ingeniería, arquitectura, urbanismo y afines, área que presenta una tasa de graduación de 29,18% en el nivel de formación universitario.

Tabla 2

Tasa de graduación 2015 por área de conocimiento para los niveles TyT agregada y Universitaria

Área de conocimiento	TyT agregada	Universitaria
Agronomía, veterinaria y afines	23,23%	24,20%
Bellas artes	21,65%	37,82%
Ciencias de la educación	19,08%	37,97%
Ciencias de la salud	38,17%	44,32%
Ciencias sociales y humanas	63,97%	33,71%
Economía, administración, contaduría y afines	25,41%	36,43%
Ingeniería, arquitectura, urbanismo y afines	18,49%	29,18%
Matemáticas y ciencias naturales	35,60%	26,25%

Fuente: SPADIES corte abril 2016. En: (M. de E. de Colombia, 2015)

Deserción en el contexto institucional

La Universidad de Cundinamarca no es ajena al problema de la deserción estudiantil. En la Figura 3 se observa el porcentaje de deserción institucional a lo largo del tiempo por período académico.

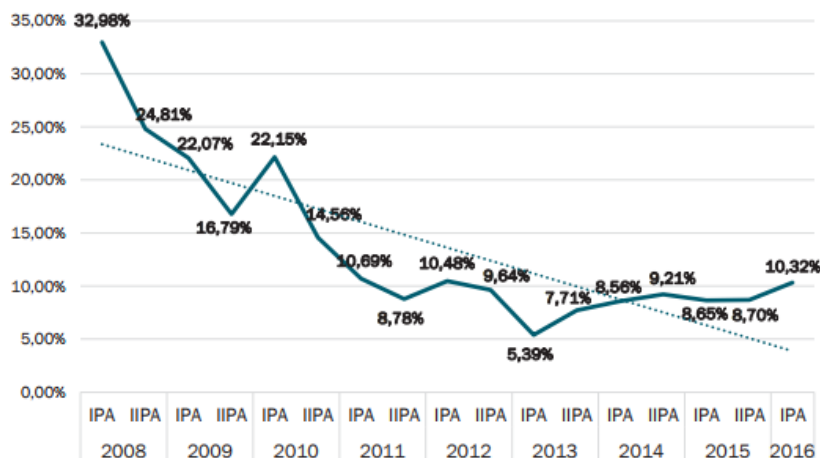


Figura 3. Porcentaje de deserción Universidad de Cundinamarca

Fuente: (Universidad de Cundinamarca, 2015)

De igual manera, en cada una de las sedes, seccionales y extensiones se presenta deserción estudiantil. En la Figura 4 se expone un gráfico de la deserción por unidad regional, donde se evidencia la existencia del fenómeno en la extensión Facatativá.

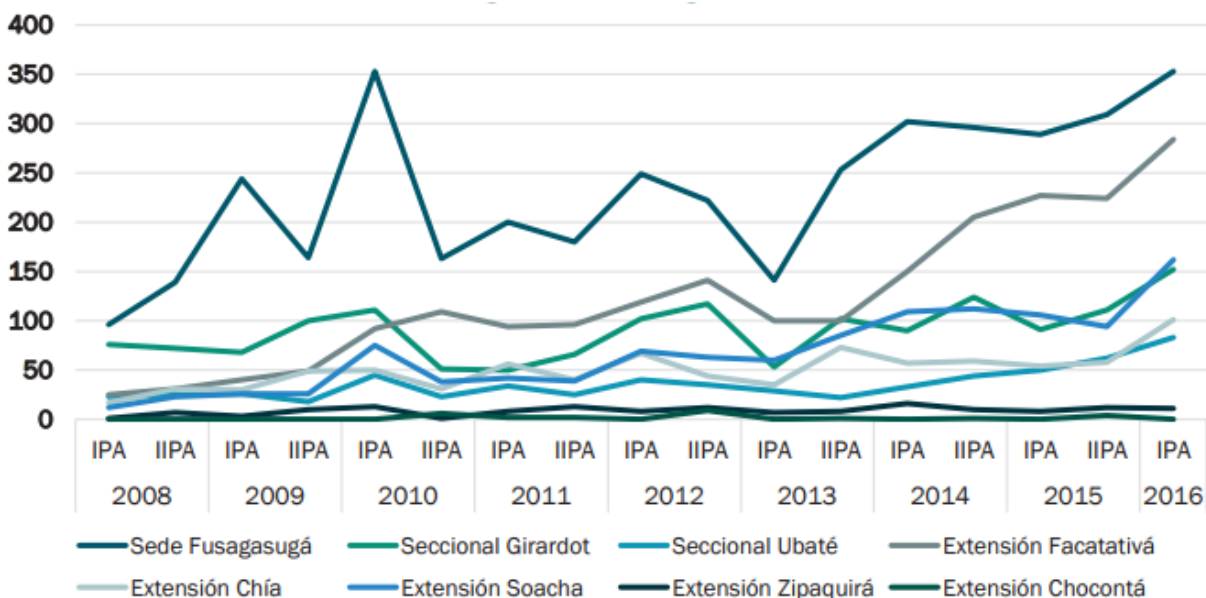


Figura 4. Deserción programas de pregrado I P.A. 2008- I P.A. 2016

Fuente: (Universidad de Cundinamarca, 2015)

En la Tabla 3 se puede observar la cantidad de desertores y el porcentaje de deserción por periodo académico para el programa de ingeniería de sistemas en la extensión Facatativá.

Tabla 3

Cantidad de desertores para el programa de ingeniería de sistemas UDEC – Facatativá

UdeC - Facatativá		Deserción IS	Total x Año	Porcentaje Deserción
2008	I P.A.	9	18	39.13%
	II P.A.	9		26.47%
2009	I P.A.	13	27	22.03%
	II P.A.	14		16.28%
2010	I P.A.	36	51	32.73%
	II P.A.	15		23.44%
2011	I P.A.	16	31	12.50%
	II P.A.	15		9.93%
2012	I P.A.	23	55	12.92%
	II P.A.	32		16.75%
2013	I P.A.	12	28	5.77%
	II P.A.	16		7.41%
2014	I P.A.	24	48	10.76%
	II P.A.	24		10.81%
2015	I P.A.	33	60	12.89%
	II P.A.	27		10.93%
2016	I P.A.	36	36	12.86%
Promedio		20.82		
Total matriculados			354	

Fuente: (Universidad de Cundinamarca, 2015). En: (Pertuz Arroyave & Chaves Sánchez, 2016)

Minería de datos

Antes del concepto de minería de datos, es preciso dejar claro otros conceptos relacionados al tema.

Datos: Los datos son en esencia números o texto que puede ser procesado en una computadora, en la actualidad las organizaciones acumulan grandes cantidades de datos en distintos formatos y en distintas bases de datos, entre las que se incluyen datos operacionales o transaccionales en las que se almacenan costos, ventas, inventarios, contabilidad, entre otros (Méndez, 2006).

Información: Los patrones, asociaciones o relaciones entre los datos proporcionan información, por ejemplo, el análisis de transacciones de un punto de venta nos puede dar información sobre qué cantidad de productos se han vendido y durante cuánto tiempo (Méndez, 2006).

Conocimiento: La información puede ser convertida en conocimiento partiendo de patrones históricos (Méndez, 2006).

Definición de minería de datos

Aunque desde un punto de vista académico el término data mining es una etapa dentro de un proceso mayor llamado extracción de conocimiento en bases de datos (Knowledge Discovery in Databases o KDD) en el entorno comercial, ambos términos se usan de manera indistinta. Lo que en verdad hace la minería de datos es reunir las ventajas de varias áreas como la Estadística, la Inteligencia Artificial, la Computación Gráfica, las Bases de Datos y el Procesamiento Masivo, principalmente usando como materia prima las bases de datos. Una definición tradicional es la siguiente: "Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos". Desde otro punto de vista se define como "la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión" (Virsedá Benito & Román Carrillo, 2010).

Áreas relacionadas con la minería de datos

La minería de datos es un campo multidisciplinar que se ha desarrollado en paralelo con otras tecnologías, por tal razón la investigación y los avances del campo se nutren de los que se producen en las siguientes áreas relacionadas (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004):

- Bases de Datos: conceptos como almacenes de datos y el procesamiento analítico en línea (OLAP) tienen gran relación con la minería de datos, pero en este caso no se trata de obtener informes avanzados a base de agregar los datos de manera predefinida, sino de extraer conocimiento novedoso y comprensible. Para el diseño de algoritmos eficientes de minería de datos son relevantes las técnicas de indización y de acceso eficiente a los datos.
- Recuperación de la información (Information retrieval): consiste en obtener información desde datos textuales y en la búsqueda por internet, donde una de las tareas típicas es encontrar documentos a partir de palabras claves, lo cual podría ser visto como un proceso de clasificación de los documentos en función de las palabras clave. Para tal fin se usan medidas de similitud entre los documentos y la consulta.
- Estadística: esta área ha proporcionado la mayoría de los conceptos, algoritmos y técnicas que se utilizan en minería de datos, por ejemplo, la media, la varianza, las distribuciones, el análisis univariante y multivariante, las regresiones y la modelización paramétrica, entre otros.
- Aprendizaje automático (Machine learning): es el área de la inteligencia artificial encargada de desarrollar algoritmos capaces de aprender y se ha convertido junto a la estadística en el corazón del análisis inteligente de datos.
- Sistemas para la toma de decisión: son sistemas informáticos que ayudan a los directivos de las organizaciones en la resolución de problemas, en la toma de decisiones y en tareas de diagnóstico.
- Visualización de datos: permite al usuario descubrir, intuir o entender patrones que serían imposibles de ver a partir de descripciones matemáticas o textuales de los resultados.

- Computación paralela y distribuida: en este tipo de sistemas el costo computacional de las tareas de minería de datos se reparte entre diferentes dispositivos. La adaptación distribuida de almacenes de datos hace posible el aumento de la escalabilidad de los algoritmos de minería de datos.
- Otras Disciplinas: según el tipo de datos a ser minados, se usan técnicas de otras disciplinas como el lenguaje natural, el análisis de imágenes, el procesamiento de señales, los gráficos por computadora, entre otras.

Procedimiento general del proceso de minería de datos

El proceso genérico de minería de datos consta de tres etapas: una etapa inicial de exploración, integración y limpieza de datos; una etapa de definición de patrones y una etapa de validación de modelos.

A. Exploración, integración y limpieza de los datos

Para empezar, se tiene el proceso de exploración de datos, que consiste en hacer uso de una metodología estructurada con el fin de descubrir y evaluar problemas apropiados, definir soluciones de tal manera que al final del proceso se tengan resultados de utilidad. Dentro de esta etapa se identifica y limita el problema, se especifica qué resultados se desean obtener y el método para implementarlos, posteriormente se preparan los datos realizando procedimientos de integración, limpieza y evaluación con el fin de tener datos completos y aptos para las siguientes etapas (Martinez, 2004).

B. Definición de patrones o construcción de modelos

En esta etapa se aplican los algoritmos y técnicas de minería de datos que se ajustan a las características de los datos disponibles (Martinez, 2004).

C. Validación y verificación de modelos

Una vez se hayan creado modelos a partir de los datos de origen, utilizando las técnicas de minería de datos a disposición, se contrastan los modelos con los datos de origen para saber en qué medida se ajustan a la realidad (Martinez, 2004).

Metodologías para la minería de datos

La metodología está conformada por procedimientos o métodos para la construcción de la evidencia empírica, por su parte, los métodos constituyen una serie de pasos que el investigador sigue en el proceso de producir una contribución al conocimiento (Sautu, 2005). Las metodologías permiten llevar a cabo el proceso de minería de datos en forma sistemática y no trivial. Algunos modelos conocidos como metodologías son en realidad un modelo de proceso: un conjunto de actividades y tareas organizadas para llevar a cabo un trabajo. La diferencia fundamental entre metodología y modelo de proceso radica en que el modelo de proceso establece qué hacer, y la metodología especifica cómo hacerlo. Una metodología no solo define las fases de un proceso sino también las tareas que deberían realizarse y cómo llevarlas a cabo (Moine et al., 2011).

Las metodologías para la gestión de proyectos de minería de datos más difundidas en la comunidad científica son KDD, SEMMA, CRISP-DM y Catalyst; sin embargo, existen otros enfoques experimentales y de menor difusión que crean híbridos con estándares de Ingeniería de Software (Moine, 2013).

A. SEMMA

SEMMA, creada por SAS Institute, fue propuesta especialmente para trabajar con el software SAS Enterprise Miner. Si bien en la comunidad científica se conoce a SEMMA como una metodología, en el sitio de la empresa SAS se aclara que éste no es el objetivo de la misma,

sino más bien la propuesta de una organización lógica de las tareas más importantes del proceso de minería de datos. Las etapas que componen la metodología son (Moine, 2013):

- **Sample (Muestreo):** En esta etapa se toma una muestra del conjunto de datos disponible, que debe ser lo suficientemente grande para contener la información relevante, y lo suficientemente pequeña como para correr el proceso rápidamente. La etapa de muestreo es opcional, aconsejable cuando el tamaño del conjunto de datos es demasiado extenso.
- **Explore (Exploración):** Consiste en explorar los datos en búsqueda de relaciones y tendencias desconocidas. Es una etapa especial para familiarizarse con los datos, y formular nuevas hipótesis a partir de su análisis.
- **Modify (Modificación):** Consiste en una etapa de preparación de los datos, donde se limpian los valores anómalos, se realiza un tratamiento de los datos faltantes, y se seleccionan, crean y modifican las variables con las que se trabajarán.
- **Model (Modelado):** Consiste en la creación del modelo que permitirá predecir las variables de respuesta a partir de las variables explicativas, utilizando algunas de las técnicas predictivas como árboles de decisión, redes neuronales, análisis discriminante o análisis de regresión.
- **Assess (Evaluación):** En esta fase se evalúa la utilidad y la exactitud de los modelos obtenidos en el proceso de minería de datos, por ejemplo, analizando la capacidad predictiva de los mismos.
- SEMMA propone que luego de la fase de evaluación, se generan nuevas hipótesis que llevan a repetir el proceso iterativamente.

B. CATALYST

Es una metodología propuesta por Dorian Pyle para el proceso de extracción de conocimiento en bases de datos. Pyle recomienda que el proceso de minería de datos siempre debería colaborar con una situación organizacional, como un problema u oportunidad. El autor aconseja no trabajar directamente con los datos sino establecer de antemano la problemática que se aborda, el personal involucrado y las expectativas y necesidades de los usuarios. Este punto resulta de gran importancia para justificar la realización del proyecto, ya que difícilmente una organización compre una herramienta si no sabe la función que cumplirá.

Para proyectos donde el problema u oportunidad de negocio no está definido, se recomienda comenzar analizando las relaciones P3TQ - Product (Producto), Place (Lugar), Price (Precio), Time (Tiempo) y Quantity (Cantidad) - que existen en la cadena de valor organizacional. Las relaciones P3TQ se refieren a tener el producto correcto, en el lugar adecuado, en el momento adecuado, en la cantidad correcta y con el precio correcto. Por su parte, la cadena de valor empresarial es un modelo teórico popularizado por Michael Porter que define las actividades de la empresa que van añadiendo valor al producto a medida que éste pasa por cada una de ellas. El foco que la metodología Catalyst propone sobre la cadena de valor, hizo que la misma sea difundida en la comunidad científica como metodología P3TQ, aunque esta no sea su denominación original (Moine, 2013).

La metodología Catalyst está formada por dos partes: Metodología para el Modelado del Negocio y Metodología para la Minería de Datos:

- Metodología para el Modelado del Negocio: En esta primera parte se proporciona una guía de pasos para modelar el problema u oportunidad de negocio que abordará el proyecto.

Pyle propone cinco situaciones o puntos de partida diferentes para el proyecto (Moine, 2013):

- Escenario 1, Datos: Se exploran los datos en búsqueda de relaciones útiles e interesantes.
- Escenario 2, Problema: Ver como la minería de datos puede colaborar en la resolución del problema.
- Escenario 3, Prospección: Diseño de proyecto para descubrir donde la minería de datos puede aportar valor en la organización.
- Escenario 4, Modelo Definido: Construir un modelo específico usando la minería de datos para una situación determinada.
- Escenario 5, Estrategia: Dada una situación estratégica, analizar si la minería de datos puede ser útil para explicar la situación actual y descubrir cuáles son las opciones para resolverla.

Cualquiera sea el escenario de partida, siempre se modela el problema que el proyecto de minería abordará, obteniendo los datos necesarios para minar y los requerimientos reales, considerando las expectativas y necesidades del usuario (Moine, 2013).

- Metodología para la Minería de Datos: Proporciona una guía de pasos para el descubrimiento de patrones y relaciones de acuerdo al problema de negocio identificado.

Se compone de las siguientes cuatro etapas (Moine, 2013):

- Preparación de los datos.
- Selección de herramientas y modelado inicial.
- Refinación del modelo.

- Implementación del modelo.

La metodología Catalyst, tanto en la parte de modelado de negocio como en la minería de datos está compuesta por una serie de pasos llamados boxes. La idea es que luego de llevar a cabo una acción, se deben evaluar los resultados y determinar cuál es el próximo paso a seguir (el siguiente box). Algunos pasos permiten al modelador elegir entre múltiples caminos dada una situación. La secuencia y la interacción entre los distintos pasos permiten una flexibilidad muy grande, y una amplia variedad de caminos posibles. Existen cuatro tipos de “boxes”: “Action Boxes”, “Discovery Boxes”, “Technique Boxes”, y “Example Boxes”. Cada tipo cumple un rol específico dentro de la metodología (Moine, 2013).

- Action boxes (AB), en las que se determina cuáles son los siguientes pasos a llevar a cabo en una determinada situación.
- Discovery boxes (DB), en las que se evalúan los resultados y posibles problemas luego de ejecutar una acción.
- Technique boxes (TB), describe cómo utilizar una determinada técnica.
- Example boxes (EB), ejemplifican la aplicación de una técnica. Existe sólo uno en toda la metodología.

C. CRISP-DM

CRISP-DM (CRoss Industry Standard Process for Data Mining) fue presentada en el año 1999 por las empresas SPSS, Daimler Chrysler y NCR. Es una metodología abierta, no está ligada a ningún producto comercial, y fue construida en base a la experiencia de sus creadores, es decir desde un enfoque práctico. La metodología está estructurada en un proceso jerárquico,

compuesto por tareas descritas en cuatro niveles diferentes de abstracción, que van desde lo general a lo específico. CRISP-DM, propone en el nivel más alto seis fases para el proceso de minería de datos: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación e implementación. La sucesión de fases, no es necesariamente rígida (Moine, 2013).

Cada fase se descompone en un conjunto de tareas genéricas (o generales) de segundo nivel. Estas tareas son genéricas ya que tratan de abarcar la mayoría de las situaciones posibles en minería de datos. A partir del tercer nivel de abstracción, se realiza un “mapeo” de las tareas genéricas definidas en el modelo a situaciones específicas. De esta forma, las tareas genéricas se traducen en tareas específicas para casos y proyectos concretos. En el cuarto nivel, encontramos las instancias de proceso, donde se describen las acciones, decisiones y resultados de un proyecto particular de minería de datos (Moine, 2013).

La metodología proporciona un modelo de referencia y una guía de usuario. El modelo de referencia presenta un resumen de las fases y tareas a llevar a cabo en cada una junto con sus salidas, describiendo lo que debería hacerse en un proyecto de minería de datos. La guía de usuario proporciona sugerencias para la ejecución de cada tarea del modelo de referencia (Moine, 2013).

En el nivel más alto de abstracción del modelo CRISP-DM, las seis fases que componen el proceso de minería de datos son (Moine, 2013):

- Comprensión del negocio: en esta fase se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, definiendo el problema de minería y el plan de trabajo.

- Comprensión de los datos: fase que consiste en la recolección de datos que se utilizarán en el proyecto y la familiarización con los mismos. En esta etapa es posible el surgimiento de las primeras hipótesis acerca de la información que podría estar oculta.
- Preparación de los datos: comprende aquellas actividades de tratamiento de los datos para construir el conjunto de datos final sobre el cual se aplicarán las técnicas de minería.
- Modelado: en esta etapa se aplican las diversas técnicas y algoritmos de minería sobre el conjunto de datos para obtener la información oculta y los patrones implícitos en ellos.
- Evaluación: fase en la que se analizan los patrones obtenidos en función de los objetivos organizacionales. En esta etapa se debería determinar si se ha omitido algún objetivo importante del negocio y si el nuevo conocimiento será implementado, es decir, si se pasará a la próxima etapa.
- Implementación: consiste en la comunicación e implementación del nuevo conocimiento, el cual debe ser representado de forma entendible para el usuario.

D. KDD

La minería de datos es un paso en el proceso de descubrimiento de conocimiento en bases de datos (KDD, del inglés Knowledge Discovery from Databases). Este proceso consta de las etapas de preparación de datos (las dos primeras fases), minería de datos, evaluación y difusión (Hernández Orallo et al., 2004):

- Fase de integración y recopilación: en esta fase se determinan las fuentes de información que son útiles y donde conseguirlas. Luego se transforman todos los datos a un formato común en almacenes de datos que consigan unificar toda la información recogida detectando y resolviendo inconsistencias.
- Fase de selección, limpieza y transformación: dado que los datos provienen de diferentes fuentes, pueden contener valores erróneos o faltantes, esto se corrige en esta fase, en donde se eliminan o corrigen datos incorrectos y se decide la estrategia a seguir con los datos incompletos, además se proyectan los datos para considerar los atributos o variables que van a ser relevantes para facilitar la tarea de minería y garantizar que los resultados de la misma sean los más útiles.
- Fase de minería de datos: en esta etapa se decide cuál es la tarea de minería de datos a realizar y se elige la técnica que se va a utilizar.
- Fase de evaluación e interpretación: en esta fase se evalúan los patrones y se analizan por los expertos, y si es necesario se vuelve a las fases anteriores para una nueva iteración.
- Fase de difusión: en esta fase se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios.

Tareas de minería de datos

En la minería de datos se pueden realizar dos tipos de tareas, las tareas predictivas y las tareas descriptivas; las predictivas permiten predecir uno o más valores para uno o más ejemplos y las descriptivas permiten descubrir reglas que muestran relaciones entre variables (Virsedo Benito & Román Carrillo, 2010).

A. Tareas predictivas

- **Clasificación:** El objetivo de la tarea es poder clasificar un dato dentro de las clases definidas del dominio que se está modelando. Como ejemplos podrían mencionarse: la clasificación de correos electrónicos como spam o no, la clasificación de medicamentos para saber cuál es el mejor para cierta enfermedad y la clasificación de clientes puntuales e impuntuales con los préstamos (Ordóñez, 2013).
- **Regresión:** El objetivo de la tarea es poder encontrar la similitud entre valores de atributos de una determinada clase de un dominio dado. Como ejemplos se pueden mencionar: la predicción del número de unidades defectuosas de una partida de productos y determinar la demanda de un producto en una tienda analizando las ventas anteriores (Ordóñez, 2013).
- **Preferencias o priorización:** Consiste en determinar a partir de dos o más ejemplos, un orden de preferencia (Ordóñez, 2013).

B. Tareas descriptivas

- **Agrupamiento o clustering:** Su objetivo es obtener grupos o conjuntos de datos en donde se incorporen elementos similares extraídos de las clases del dominio dado. Como ejemplo se puede mencionar la agrupación de clientes por segmentos según sus perfiles para poder estudiar el comportamiento de cada grupo ante determinados productos (Ordóñez, 2013).
- **Asociación:** El objetivo de esta tarea es poder describir las relaciones que existen entre los valores de los atributos de un determinado ejemplo de un dominio establecido. Por ejemplo, se puede realizar un análisis de las compras de varios clientes con el fin de mejorar la ubicación de los productos en un supermercado (Ordóñez, 2013).

- **Correlación:** Su objetivo es ver si dos o más atributos, dados los elementos de un conjunto, están correlacionados linealmente o relacionados de algún otro modo mediante un análisis de varianza coeficiente de correlación lineal de los datos. Por ejemplo, se puede analizar los factores que influyen en la asistencia de pacientes a un centro de salud (Ordóñez, 2013).

Técnicas de minería de datos

Debido a que la mayoría de las tareas de minería de datos pertenecen al campo del aprendizaje inductivo, una tarea puede tener muchas técnicas diferentes para resolverla y una misma técnica puede resolver varios tipos de tareas, a continuación, se exponen los tipos de técnicas existentes (Hernández Orallo et al., 2004):

- **Técnicas algebraicas y estadísticas:** se basan en fórmulas algebraicas, funciones lineales y no lineales, distribuciones o valores agregados estadísticos tales como medias, varianzas, correlaciones, etc. Frecuentemente son técnicas que estiman coeficientes o parámetros a partir de un modelo predeterminado, conocidas como técnicas paramétricas. No obstante, existen técnicas de modelización estadística no paramétrica. Ejemplos de algoritmos de esta técnica son la regresión lineal, la regresión logarítmica, la regresión logística y los discriminantes no lineales.
- **Técnicas bayesianas:** estiman la probabilidad de pertenencia a una clase o grupo utilizando el teorema de Bayes. Las redes bayesianas generalizan las topologías de las interacciones probabilísticas entre variables y permiten representar gráficamente dichas interacciones.
- **Técnicas basadas en conteos de frecuencias y tablas de contingencia:** cuentan la frecuencia en la que dos o más sucesos se presentan conjuntamente. Cuando el conjunto de sucesos posibles es muy grande existen algoritmos que van comenzando por pares de sucesos e

incrementando los conjuntos sólo en aquellos casos que las frecuencias conjuntas superen un cierto umbral.

- Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas: además de tener representaciones en forma de reglas se basan en dos tipos de algoritmos: algoritmos de tipo divide y vencerás, y los algoritmos denominados separa y vencerás.
- Técnicas relacionales, declarativas y estructurales: representan los modelos mediante lenguajes declarativos como lenguajes lógicos, funcionales o lógico-funcionales.
- Técnicas basadas en redes neuronales artificiales: aprenden un modelo mediante el entrenamiento de los pesos que conectan un conjunto de nodos o neuronas.
- Técnicas basadas en núcleo y máquinas de soporte vectorial: se basan en transformaciones que pueden aumentar la dimensionalidad para maximizar el margen entre los grupos o las clases formadas. Dichas transformaciones se denominan núcleos o kernels.
- Técnicas estocásticas y difusas: son técnicas en las que los componentes aleatorios son fundamentales como en el recocido simulado (simulated annealing) y técnicas como los métodos evolutivos, genéticos y la utilización de funciones de pertenencia difusas. Estas técnicas, junto con las redes neuronales, forman lo que se denomina computación flexible o soft computing.
- Técnicas basadas en casos, en densidad o distancia: se basan en distancias al resto de elementos, como los vecinos más próximos o estimación de funciones de densidad. Otros algoritmos muy conocidos son los jerárquicos y los no jerárquicos como K medias.
- Además de las técnicas anteriores, existen híbridos y métodos por combinación.

En la Tabla 4 se muestra una correspondencia entre tareas de minería de datos y algunos algoritmos que pueden abordarlas.

Tabla 4
Algoritmos y tareas de minería de datos

Nombre	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Reglas de asociación	Correlaciones / Factorizaciones
Redes neuronales	✓	✓	✓		
Arboles de decisión ID3, C4.5, C5.0	✓				
Arboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión Logística	✓			✓	
Kmeans			✓		
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis factorial y de comp. Ppales.					✓
Twostep, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores de soporte	✓	✓	✓		
CN2 rules (cobertura)	✓			✓	
Análisis discriminante multivariante	✓				

Fuente: (Hernández Orallo et al., 2004). En: (Pertuz Arroyave & Chaves Sánchez, 2016)

Herramientas para minería de datos

Las herramientas de minería de datos facilitan el desarrollo de modelos para la extracción de conocimiento de un dominio establecido, dichas herramientas contienen los algoritmos específicos para la aplicación de técnicas de minería de datos. A continuación se exponen algunas herramientas de minería de datos (Ordóñez, 2013):

- **Spss clementine:** Es uno de los sistemas más conocidos, se caracteriza por su acceso a datos con archivos ASCII, por el procesamiento de datos, la aplicación de técnicas de aprendizaje como redes neuronales, y además incorpora técnicas de evaluación de modelos y visualización de resultados como histogramas y diagramas de dispersión.
- **WEKA (Waikato environment for Knowledge analysis):** Es una herramienta visual de libre distribución desarrollada por los investigadores de la universidad de Waikato en Nueva Zelanda. Sus características destacadas son el acceso de los datos desde archivos ARFF, su preprocesador de datos, la visualización del entorno y la aplicación de técnicas de aprendizaje como redes neuronales y arboles de decisión.
- **Kepler:** Es una herramienta comercial distribuida por Dialogis. Posee múltiples modelos de análisis como redes neuronales, regresión no lineal y aplicaciones estadísticas. Permite el pre procesamiento de datos y la manipulación de la representación gráfica de los modelos obtenidos.
- **ODMS (Oracle data mining suite):** Diseñada sobre una arquitectura cliente servidor; ofrece una gran versatilidad en cuanto al acceso a grandes volúmenes de información. Se caracteriza por su acceso de datos desde diversos formatos y motores de bases de datos relacionales. Además, brinda herramientas de visualización para resultados estadísticos e importación de datos en Excel, Word o Power point.

- Dbminer: Está concebido para la extracción del conocimiento de bases de datos relacionales, almacenes de datos y web con la utilización de técnicas como reglas de asociación, reglas difusas e incorpora tareas de clasificación.
- Rapid miner: Es una herramienta de aprendizaje automático que incluye operaciones para importación y pre procesamiento de datos, validación de modelos y permite la aplicación de técnicas como redes neuronales y reglas de asociación.
- DB2 intelligent miner: Herramienta comercial distribuida por IBM que permite procesar grandes cantidades de datos.
- SAS Enterprise miner: Posee una arquitectura distribuida y una potente interfaz de usuario, permite realizar tareas de pre procesamiento, tratamiento estadístico, filtros, tareas de muestreo y además una visualización de resultados a través de gráficos, diagramas e informes en formato HTML.
- Statistica data miner: Es una herramienta con un sistema visual, tiene soporte de bases de datos, pre procesamiento de datos, modelos de análisis, visualización de datos y múltiples diagramas.
- Cart: Orientada a tareas de clasificación o regresión de minería de datos. Se destaca principalmente por su accesibilidad, capacidad de visualización e información estadística relativa al modelo.

Áreas de aplicación

Los negocios de la distribución y la publicidad dirigida han sido tradicionalmente las áreas en las que más se han empleado los métodos de minería, ya que han permitido reducir costos o aumentar la receptividad de ofertas. Pero estas no son las únicas áreas donde se puede aplicar, se pueden encontrar ejemplos en todo tipo de aplicaciones: financieras, seguros, científicas, políticas,

económicas, sanitarias o demográficas, educaciones policiales, procesos industriales, entre otras (Hernández Orallo et al., 2004).

A continuación, se listan algunas aplicaciones de la minería de datos en distintas áreas (Hernández Orallo et al., 2004):

- Aplicaciones financieras y banca: obtención de patrones de uso fraudulento de tarjetas de crédito, análisis de riesgos en créditos, cálculo de correlaciones entre indicadores financieros.
- Análisis de mercado y comercio: Evaluación de campañas publicitarias, segmentación de clientes estimación de inventarios.
- Seguros y salud privada: Identificación de patrones de comportamiento para clientes con riesgo, identificación de comportamientos fraudulentos.
- Educación: Selección o captación de estudiantes, detección de abandonos y de fracaso.
- Procesos industriales: modelos de calidad, predicción de fallos y accidentes, extracción de modelos de producción.
- Medicina: Diagnóstico de enfermedades, gestión hospitalaria y asistencial.
- Biología: análisis de secuencias de genes, modelos de calidad de aguas, indicadores ecológicos, clasificación de cuerpos celestes.
- Telecomunicaciones: Establecimiento de patrones de llamadas, modelos de carga en redes, detección de fraude.
- Otras áreas: Correo electrónico, selección de empleados, análisis web, turismo, modelos de tráfico, planificación de eventos, diseño de campañas políticas, entre otras.

Minería de datos educacional

La minería de datos aplicada a la educación o EDM es una disciplina emergente que desarrolla y aplica métodos de minería de datos sobre datos que vienen de entornos educativos, y los usa para entender mejor a los estudiantes y los entornos en los que aprenden. Además del término EDM existen también otros términos similares que son empleados por áreas muy afines o relacionadas con EDM como son (Márquez Vera, 2015):

- Análisis del aprendizaje o Learning Analytics: consiste en la medida, recolección, análisis e informe de datos sobre estudiantes/aprendices y su contexto, con el propósito de comprender y optimizar el aprendizaje y el entorno donde ocurre.
- Análisis académico o Academic Analytics: consiste en la aplicación de técnicas estadísticas y de minería a datos institucionales para producir inteligencia de empresa y soluciones a universidades y administradores.

EDM tiene por objetivos fundamentales mejorar y guiar el proceso de aprendizaje de los estudiantes, además realizar investigación que permita profundizar en el entendimiento de los fenómenos educativos. Estas metas son en ocasiones difíciles de cuantificar y requieren de un propio conjunto de técnicas de medición (Márquez Vera, 2015).

El conocimiento obtenido mediante EDM puede ser usado por diferentes actores del proceso educativo, aunque en consideraciones iniciales se veía que iba dirigido para los estudiantes y los profesores, actualmente hay más interesados y con diferentes objetivos, como los diseñadores de cursos, los investigadores de los procesos educativos, el personal directivo-administrativo de las instituciones, entre otros (Márquez Vera, 2015).

Entre los campos del ámbito educativo que pueden ser tratados usando la minería de datos están (Márquez Vera, 2015):

- Estudiar sobre el soporte pedagógico que aporta el software de aprendizaje.
- La investigación científica sobre el aprendizaje y los alumnos.
- Las aplicaciones para la evaluación del desempeño del aprendizaje de los alumnos.
- Las aplicaciones para que se dé la retroalimentación alumno-profesor en los cursos en línea.
- Las aplicaciones para detectar a alumnos con comportamientos atípicos.
- Análisis y visualización de la información.
- Recomendaciones para los estudiantes.
- Adaptación de modelos educativos para los estudiantes.
- Detección de comportamientos indeseables en los alumnos.
- Análisis de redes sociales.
- Desarrollo de mapas conceptuales.
- Construcción de cursos para la web.
- Planeación y programación de cursos.
- Predicción del rendimiento académico de los estudiantes.

Lenguaje de programación python

El resultado de la presente investigación, un aplicativo web para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas de la Universidad de Cundinamarca en la extensión de Facatativá utilizando modelos predictivos de minería de datos se desarrolla con el lenguaje de programación Python, la librería Scikit-learn y el framework web Flask.

Python es un lenguaje de programación que se está estableciendo como uno de los lenguajes de programación más populares para la computación científica. Es un lenguaje de programación interpretado que dispone de un conjunto maduro de bibliotecas. Debido a sus características es una opción atractiva para el desarrollo algorítmico y el análisis exploratorio de datos (Pedregosa et al., 2012).

La historia del lenguaje de programación Python se remonta a finales de la década de 1980. Python fue concebido a finales de los años ochenta y su puesta en práctica se inició en diciembre de 1989 por Guido van Rossum en el Centrum Wiskunde & Informatica en los Países Bajos como sucesor del lenguaje de programación ABC. Python es un lenguaje de programación de uso general y de alto nivel, su diseño enfatiza en la legibilidad del código y su sintaxis permite a los programadores expresar conceptos en menos líneas de código que en lenguajes como C++ o Java (Klein, 2015).

Python soporta múltiples paradigmas de programación, incluyendo el paradigma orientados a objetos, la programación funcional, procedimental, programación orientada a aspectos, y con el uso de extensiones, el diseño por contrato y la programación lógica. Cuenta con un sistema de tipos de variables dinámico y memoria automática. Los intérpretes de Python están disponibles para su instalación en múltiples sistemas operativos, permitiendo la ejecución de código Python en una amplia variedad de sistemas (Klein, 2015).

Scikit-learn

Scikit-learn aprovecha el entorno de Python para proporcionar implementaciones de vanguardia de algoritmos de aprendizaje de la máquina bien conocidos, manteniendo una interfaz fácil de usar integrada con el lenguaje Python. Esto responde a la creciente necesidad de análisis de datos por parte de personas no especialistas en el software y la industria web, así como en

campos fuera de la informática, tales como la biología o la física. Scikit-learn difiere de otras herramientas de aprendizaje de máquina por diversas razones (Pedregosa et al., 2012):

- Se distribuye bajo la licencia BSD.
- Incorpora código compilado para aumentar la eficiencia en el uso de recursos de máquina.
- Depende sólo de las librerías Numpy y Scipy para facilitar su distribución.
- Se centra en la programación imperativa.

Flask

Flask es un marco de desarrollo (framework) web pequeño según la mayoría de los estándares, lo suficientemente pequeño como para ser llamado un microframework, es por eso que se puede llegar a entender su código fuente rápidamente. Por ser pequeño no significa que haga menos que otros frameworks. Flask fue diseñado para ser extensible desde el principio; proporciona un núcleo sólido con los servicios básicos, mientras que el resto lo proporcionan extensiones. Con la ventaja de que se pueden elegir justamente las extensiones que se necesiten (Grinberg, 2015).

Flask tiene dos dependencias principales. Los subsistemas de rutas, depuración y una interfaz Web Server Gateway (WSGI) que proceden del paquete Werkzeug y el motor de plantillas Jinja2 (Grinberg, 2015).

No hay soporte nativo en Flask para acceder a bases de datos, validar formularios web, autenticación de usuarios u otras tareas de alto nivel. Estos y muchos otros servicios clave de aplicaciones web están disponibles a través de extensiones que se integran con los paquetes principales. Como desarrollador, se tiene el poder de seleccionar las extensiones que funcionan mejor para un proyecto o incluso escribir extensiones personalizadas, en contraste con frameworks

más robustos, donde la mayoría de las opciones ya vienen incluidas y son difíciles o a veces imposibles de cambiar (Grinberg, 2015). Debido a su extensibilidad, Flask se puede integrar fácilmente con librerías para el análisis científico de datos como Scikit-learn.

Arquitectura orientada a recursos

La plataforma web MIDUDEC resultado del presente proyecto de investigación se diseña y construye en una arquitectura orientada a recursos con la tecnología RESTful debido a las ventajas que presenta esta arquitectura en cuanto a simplicidad, reutilización de componentes y escalabilidad.

Una arquitectura orientada a recursos (ROA) es un patrón de diseño arquitectónico y su correspondiente paradigma de programación proporciona una metodología conceptual y herramientas de desarrollo para crear arquitecturas distribuidas. Las arquitecturas distribuidas consisten en componentes que los clientes y otros componentes pueden acceder a través de una red por medio de una interfaz (Lucchi, Millot, & Elfers, 2008).

Las ROA se basan en el concepto de recurso; cada recurso es un componente distribuido directamente accesible que se maneja a través de una interfaz estándar. La tecnología de desarrollo de transferencia de estado representacional (REST) permite la creación de ROA (Lucchi et al., 2008).

Los conceptos fundamentales de ROA son los siguientes (Lucchi et al., 2008):

- **Recurso:** Cualquier cosa que es lo suficientemente importante para ser referenciada como una cosa en sí.
- **Nombre de recurso:** Identificación única de un recurso.
- **Representación de un recurso:** Información útil sobre el estado actual de un recurso.

- Enlace de recurso: Enlace a otra representación del mismo o de otro recurso.
- Interfaz de recurso: Interfaz uniforme para acceder a un recurso y manipular su estado.

Además de ROA existe la arquitectura orientada a servicios (SOA), en SOA todos los componentes distribuidos son servicios. Un servicio proporciona funciones de negocio a su consumidor y se define como: distintas partes de la funcionalidad que proporciona una entidad a través de interfaces. En términos informáticos, un servicio es una aplicación que proporciona información y/o funcionalidad a otras aplicaciones. Los servicios son típicamente aplicaciones que se ejecutan en servidores e interactúan con otras aplicaciones a través de una interfaz (roa). Existen tres arquitecturas comunes para servicios web: RESTful orientado a recursos, llamada a procedimiento remoto (RPC) e híbrido REST-RPC (Richardson & Ruby, 2007).

- RESTful: En arquitecturas RESTful, las operaciones son manejadas con los métodos HTTP y la información de alcance se introduce en el identificador de recursos uniforme (URI).
- RPC: Un servicio web de estilo RPC acepta un sobre lleno de datos de su cliente y envía un sobre similar como respuesta. El método y la información de alcance se mantienen dentro del sobre o en pegatinas aplicadas al sobre. HTTP y SOAP son formatos de sobre populares para RPC. Un servicio de estilo RPC típicamente expone un URI para cada procesador de documentos que puede abrir sobres y transformarlos en comandos de software.
- Híbrido REST-RPC: Un híbrido REST-RPC se puede hallar en un servicio de estilo RPC que utiliza HTTP como su formato de sobre y cuando tanto el método como la información de alcance se encuentran en la parte URI de la solicitud HTTP.

Principios de una arquitectura orientada a recursos

Richardson & Ruby, (2007) mencionan los siguientes principios y sugerencias para el diseño de una interfaz de programación de aplicaciones (API) basada en una arquitectura orientada a recursos:

- Un recurso debe tener por lo menos un URI. URI es el nombre y la dirección de un recurso.
- Los URIs deberían tener una estructura uniforme y predecible para los recursos.
- Un URI debe estar designado exactamente a un sólo recurso.
- Cualquier recurso puede ser solicitado siempre con la misma dirección, por lo cual, cualquier información interesante que el servidor puede proporcionar debe ser expuesta como un recurso.
- Cada solicitud HTTP ocurre en completo aislamiento. Cuando el cliente realiza una solicitud HTTP incluye toda la información necesaria para que el servidor responda esa solicitud. El servidor nunca se basa en información de solicitudes anteriores, si esa información era importante, el cliente la envía otra vez en una nueva petición.
- Proporcionar un URI diferente para cada representación de un mismo recurso, una alternativa es la negociación de contenido entre un servidor y un navegador web por medio de cabeceras HTTP.
- Se debe procurar enviar tanta información como sea posible en el URI y no en las cabeceras HTTP.
- Los recursos deben vincularse entre sí en sus representaciones.
- En toda la Web, sólo hay algunas cosas básicas que puede hacer con un recurso. HTTP proporciona cuatro métodos básicos para las cuatro operaciones más comunes GET, POST, PUT y DELETE.

- En las peticiones para crear y modificar un recurso se usa un cuerpo de la entidad para enviar una representación al servidor. Los datos y el formato del cuerpo de la entidad dependen del servicio.
- Existen métodos HTTP no muy comunes que forman parte de una interfaz uniforme, dichos métodos tienen los siguientes usos:
 - Recuperar una representación sólo de metadatos: HTTP HEAD.
 - Comprobar qué métodos HTTP soporta un recurso en particular: HTTP OPTIONS.
- Las operaciones HTTP GET, HEAD, PUT y DELETE sobre un recurso son idempotentes, es decir, si se realiza varias veces la misma operación con los mismos parámetros se debe obtener el mismo resultado.
- Se debe evitar realizar operaciones que alteren recursos con el método HTTP GET.
- La arquitectura REST especifica que se debe usar una interfaz uniforme, pero no indica qué interfaz usar. Sin embargo, es conveniente el uso de los métodos estándar HTTP. Aunque el uso de WebDAV o de sólo HTTP GET con POST sobrecargado son interfaces uniformes, estas interfaces generan incompatibilidades con servicios RESTful, servicios que si cumplen con todos los principios de la arquitectura REST.

RESTful

REST define un conjunto de principios arquitectónicos mediante los cuales puede diseñar servicios Web que se centren en los recursos de un sistema, incluyendo cómo los estados de recursos se direccionan y se transfieren a través de HTTP por una amplia gama de clientes escritos en diferentes lenguajes de programación. Si se mide por el número de servicios web que lo utilizan, REST ha surgido en los últimos años como un modelo de diseño predominante de servicios web.

De hecho, REST ha tenido un impacto tan grande en la web que ha desplazado principalmente a SOAP y WSDL porque es considerablemente más simple de usar (Rodríguez, 2008).

REST no atrajo mucha atención cuando fue introducido por primera vez en el año 2000 por Roy Fielding en la Universidad Irvine de California en su disertación académica: "Architectural Styles and the Design of Network-based Software Architectures" que analiza un conjunto de principios de arquitectura de software que utilizan la web como una plataforma para la computación distribuida (Rodríguez, 2008).

Métodos HTTP

Una de las características clave de un servicio web RESTful es el uso explícito de los métodos HTTP de una manera consistente con la definición del protocolo. Éste principio básico de diseño REST establece un mapeo uno a uno entre las operaciones crear, leer, actualizar y eliminar (CRUD) y métodos HTTP (Rodríguez, 2008).

El mapeo usa los métodos HTTP de la siguiente manera (Richardson & Ruby, 2007):

- Recuperar una representación de un recurso: HTTP GET.
- Crear un nuevo recurso: HTTP PUT para una nueva URI, HTTP POST para un URI existente.
- Modificar un recurso: HTTP PUT para un URI existente.
- Eliminar un recurso existente: HTTP DELETE.

El uso de sustantivos en URIs en lugar de verbos ayuda a seguir las pautas de REST para el uso explícito de métodos HTTP. En un servicio web RESTful, los verbos POST, GET, PUT y DELETE ya están definidos por el protocolo. Para mantener la interfaz generalizada y permitir que los clientes sean explícitos sobre las operaciones que invocan, el servicio web no debe definir

más verbos o procedimientos remotos. Este principio de diseño general también se aplica al cuerpo de una petición HTTP, que se destina para transferir el estado de un recurso y no para llevar el nombre de un método remoto o procedimiento remoto que se invocará (Rodríguez, 2008).

Permanencia sin estado

Los servicios web REST necesitan escalarse para satisfacer demandas cada vez más exigentes. Arreglos de servidores con capacidades de balanceo de carga y conmutación por error, proxies y gateways son organizados de manera que forman una topología de servicio, la cual permite que un servidor redireccione a otro cuando sea necesario para disminuir el tiempo de respuesta global de una llamada de servicio web. El uso de servidores intermediarios para mejorar la escalabilidad requiere que los clientes de un servicio web REST envíen solicitudes independientes y completas, es decir, enviar solicitudes que incluyan todos los datos necesarios para que los componentes de los servidores intermediarios puedan reenviar, enrutar y equilibrar la carga sin mantener localmente un estado entre las solicitudes (Rodríguez, 2008).

Una solicitud completa e independiente no requiere que el servidor, mientras procesa la solicitud, necesite cualquier tipo de contexto o estado de la aplicación. Una aplicación de servicio web REST o un cliente incluye dentro de los encabezados y el cuerpo HTTP de una solicitud todos los parámetros, contexto y datos necesarios por el componente del lado del servidor para generar una respuesta. En este sentido, permanecer sin estado mejora el rendimiento de los servicios web y simplifica el diseño y la implementación de componentes porque la ausencia de estado en el servidor elimina la necesidad de sincronizar los datos de sesión con una aplicación externa (Rodríguez, 2008).

URIs estructuradas

Los URIs de servicios web REST deben ser intuitivos hasta el punto de ser fáciles de adivinar. Se debe pensar en un URI como una especie de interfaz autodocumentada que requiere poca o ninguna explicación o referencia de un desarrollador para entender lo que apunta y obtener recursos relacionados. Con este fin, la estructura de un URI debe ser sencilla, predecible y fácil de entender. Una forma de lograr este nivel de usabilidad es definir URIs con estructura de directorio. Este tipo de URI es jerárquico, originado de un único camino, y la ramificación éste en sub-rutas expone los distintos servicios (Rodríguez, 2008).

Algunas directrices adicionales para tomar nota mientras se piensa en la estructura de los URIs para un servicio web RESTful son (Rodríguez, 2008):

- Ocultar las extensiones de archivo de la tecnología de secuencias de comandos del servidor (.jsp, .php, .asp), si las hay, para que se pueda cambiar de lenguaje del lado de servidor sin cambiar los URIs.
- Mantener todo en minúsculas.
- Sustituir los espacios con guiones o guiones bajos.
- Evitar las cadenas de consulta (query strings) tanto como se pueda.
- En lugar de usar el código 404 Not Found si el URI de la solicitud está incompleto, siempre proporcionar una página o recurso predeterminado como respuesta.
- Los URIs deben ser estáticos para que cuando se realicen cambios en los recursos o en la implementación del servicio, el enlace permanezca igual.
- Es importante que las relaciones entre recursos que se codifican en los URIs sean independientes de la forma en que se almacenan.

Lenguaje de representación de recursos

Una representación de un recurso típicamente refleja el estado actual de un recurso y sus atributos en el momento en que una aplicación cliente lo solicita. Las representaciones de recursos en este sentido son simples instantáneas en el tiempo. Esto podría ser tan simple como una representación de un registro en una base de datos que consiste en una correlación entre nombres de columnas y etiquetas XML, donde los valores de elemento en el XML contienen los valores de fila; o si el sistema tiene un modelo de datos, entonces, de acuerdo con esta definición, una representación de recursos es una instantánea de los atributos de uno de los objetos en el modelo de datos de un sistema (Rodríguez, 2008).

Para dar a las aplicaciones cliente la capacidad de solicitar un tipo de contenido específico que sea más adecuado para ellos, se debe utilizar el encabezado HTTP Accept, donde el valor del encabezado es un tipo MIME. Son comunes los siguientes tipos MIME (Rodríguez, 2008):

- JSON: application/json
- XML: application/xml
- XHTML: application/xhtml+xml

Usar el encabezado Accept permite que el servicio sea utilizado por una variedad de clientes escritos en diferentes lenguajes de programación que se ejecutan en diferentes plataformas y dispositivos. El uso de tipos MIME y el encabezado HTTP Accept es un mecanismo conocido como negociación de contenido, que permite a los clientes elegir qué formato de datos es adecuado para ellos y minimiza el acoplamiento de datos entre el servicio y las aplicaciones que lo utilizan (Rodríguez, 2008).

RESTful con minería de datos

Una forma de mejorar la accesibilidad e interoperabilidad de las herramientas de minería de datos es proporcionarles una interfaz abstracta a través de servicios web. Las arquitecturas orientadas a recursos con la tecnología REST parecen adecuadas para esta tarea, dada la alineación natural de la filosofía de diseño de REST con el deseo de ocultar detalles específicos de la implementación (Reid et al., 2016).

La arquitectura protocolos y estructuras para la inferencia (PSI) sirve para la implementación de una interfaz RESTful con minería de datos. En la arquitectura de PSI, los conjuntos de datos, conocidos como relaciones (terminología tomada de las bases de datos relacionales), los atributos de esas relaciones, las transformaciones que pueden aplicarse a esos datos (y a las predicciones), aprendedores y predictores, son todos recursos que pueden ser compuestos para realizar diferentes actividades de inferencia. Cada recurso es identificado por su URI y las interacciones se manejan a través de los métodos HTTP GET, POST y DELETE. Las representaciones de los recursos, los mensajes de solicitud, respuesta, los valores de los atributos y de las predicciones están representados en JSON (Reid et al., 2016).

Los proveedores de servicios RESTful son libres de ofrecer cualquier subconjunto de recursos PSI que se adapte a sus propósitos. Por ejemplo, un servicio puede proporcionar datos a través de recursos de relaciones y atributos, mientras que otro servicio podría ofrecer algoritmos de aprendizaje y los predictores que se producen. Esta flexibilidad permite crear soluciones distribuidas de aprendizaje de máquina (Reid et al., 2016).

Un servicio PSI tiene una URI de entrada publicada única, cuya representación JSON incluye enlaces a colecciones de relaciones, aprendedores, predictores, transformadores y esquemas proporcionados por ese servicio. Cada colección tiene la misma representación, que

incluye una serie de enlaces a los recursos que contiene. De esta manera los recursos de un servicio PSI forman naturalmente una jerarquía (Reid et al., 2016).

Un servicio PSI se conforma de los siguientes tipos de recurso (Reid et al., 2016):

- Esquemas (schemas): Es conveniente proveer esquemas para indicar la estructura adecuada de los datos que consume un recurso.
- Relaciones (relations): Recursos donde se representan los conjuntos de datos (datasets), son colecciones de instancias que comparten los mismos atributos.
- Transformadores (transformers): Un transformador acondiciona datos para otros recursos. Por ejemplo, preparar datos para el aprendizaje generando atributos adicionales o convirtiendo valores en tipos de datos apropiados.
- Aprendedores (learners): Son recursos que generan predictores a partir de las relaciones tomadas como datos de entrenamiento.
- Predictores (predictors): Un predictor es una especie de transformador construido por un aprendedor.

Marco legal

En esta sección se aborda la legislación vigente que influye al presente proyecto. Debido al contexto regional y por ende nacional de la presente investigación, los aspectos legales relacionados a la minería de datos y a la herramienta de software de apoyo al proceso, plantean dos perspectivas legales a ser consideradas: el ámbito de la privacidad y protección de la información y los datos personales, y el ámbito de la protección de los derechos de autor y la propiedad intelectual (Pertuz Arroyave & Chaves Sánchez, 2016).

A continuación, se describen las normas y leyes colombianas que atañen a la presente investigación siguiendo la jerarquía legal planteada en la pirámide de Kelsen:

El artículo 15 de la constitución política de Colombia de 1991 establece los principios fundamentales para la recolección, tratamiento y circulación de datos personales, procedimientos que deben respetar la libertad y demás garantías consagradas en la constitución (Corte constitucional, 1991).

La ley 23 de 1982 dicta las disposiciones sobre protección de derechos de autor para obras literarias, científicas y artísticas (Congreso de la República de Colombia, 1982).

La ley 1273 de 2009 tipifica los delitos que conciernen a la protección de la información y de los datos en sistemas que utilizan las tecnologías de la información y las comunicaciones (Congreso de la República de Colombia, 2009).

En la ley estatutaria 1266 de 2008 se dictan las disposiciones generales del habeas data, se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones (Congreso de la República de Colombia, 2015).

La ley 1581 de 2012 tiene por objeto desarrollar el derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos o archivos, y los demás derechos, libertades y garantías constitucionales a que se refiere el artículo 15 de la constitución política; así como el derecho a la información consagrado en el artículo 20 de la misma (M. de comercio industria y turismo de Colombia, 2013).

Documentación del software

Plan de proyecto

En la Tabla 5 se relacionan las actividades planeadas para la realización de la presente investigación con sus respectivos responsables, además se especifica la duración de las tareas y los hitos en el desarrollo del proyecto.

Tabla 5
Cronograma de actividades

Actividad	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Responsable de la Actividad
Aprobación del formato MINr008	x						CIT
Levantamiento de requerimientos	x	x	x				Diego Gómez
Modelamiento de la plataforma web	x	x	x				Diego Gómez
Entrega y realimentación del modelamiento		x	x				Diego Gómez
Desarrollo investigativo y codificación de la plataforma			x	x	x		Diego Gómez
Transferencia y divulgación				x	x	x	Diego Gómez
Elaboración de documentación de la plataforma web				x	x	x	Diego Gómez
Pruebas funcionales de la aplicación				x	x	x	Diego Gómez
Realización del libro final y de artículos científicos				x	x	x	Diego Gómez
Entrega y sustentación final del proyecto						x	Diego Gómez

Fuente: Autor

Estimación de recursos

En la Tabla 6 se hace un resumen por rubros de los recursos utilizados en el presente proyecto y se detalla la cantidad de dinero en pesos colombianos solicitado a la Universidad de Cundinamarca y/o a otras entidades.

Tabla 6
Resumen de recursos por rubros

Rubros	Solicitado en efectivo a UDEC	Contrapartida en especie		Total
		UDEC	Otras Entidades	
PERSONAL	\$ 0	\$ 0	\$ 0	\$ 0
EQUIPOS	\$ 0	\$ 0	\$ 0	\$ 0
MATERIALES E	\$ 0	\$ 0	\$ 0	\$ 0
INSUMOS				
SERVICIOS	\$ 0	\$ 0	\$ 0	\$ 0
TECNOLOGICOS				
VIAJES	\$ 0	\$ 0	\$ 0	\$ 0
OTROS	\$ 0	\$ 0	\$ 0	\$ 0
TOTALES	\$ 0	\$ 0	\$ 0	\$ 0

Determinación de requerimientos (fase 2)

Tabla 7
Requerimientos funcionales

Identificación del requerimiento	RF01
Nombre del requerimiento	Autenticar el acceso del usuario de la aplicación.
Descripción del requerimiento	El sistema permitirá validar que el acceso a la aplicación sea por parte de un usuario autorizado, para ello se le pide una dirección de correo electrónico y una contraseña. En el caso de la interfaz REST, al realizar la autenticación se genera un token de acceso que se verificará para validar las peticiones.
Identificación del requerimiento	RF02
Nombre del requerimiento	Registro de usuarios.
Descripción del requerimiento	El sistema permitirá a un usuario autenticado registrar otros usuarios en la aplicación.
Identificación del requerimiento	RF03
Nombre del requerimiento	Actualizar datos de usuario.
Descripción del requerimiento	El sistema le permitirá a un usuario modificar sus datos de acceso a la aplicación.
Identificación del requerimiento	RF04
Nombre del requerimiento	Eliminar usuarios.

Descripción del requerimiento	El sistema le permitirá a un usuario autenticado eliminar cualquier cuenta de usuario incluyendo su propia cuenta.
Identificación del requerimiento	RF05
Nombre del requerimiento	Consultar usuarios.
Descripción del requerimiento	El sistema le permitirá a un usuario autenticado consultar los demás usuarios registrados en la aplicación.
Identificación del requerimiento	RF06
Nombre del requerimiento	Recuperar acceso.
Descripción del requerimiento	Esta actividad permite que un usuario envíe sus datos de acceso a su cuenta de correo electrónico.
Identificación del requerimiento	RF07
Nombre del requerimiento	Adición de datos de un aspirante o estudiante.
Descripción del requerimiento	El sistema permitirá al usuario registrar de manera individual en la plataforma de selección con minería de datos a cada uno de los aspirantes o estudiantes del programa de ingeniería de sistemas de la universidad de Cundinamarca en la extensión de Facatativá.
Identificación del requerimiento	RF08
Nombre del requerimiento	Actualizar los datos de un aspirante o estudiante.
Descripción del requerimiento	El sistema permitirá al usuario modificar datos del perfil de un aspirante o estudiante activo registrado con anterioridad.
Identificación del requerimiento	RF09
Nombre del requerimiento	Borrar un aspirante o estudiante.
Descripción del requerimiento	El sistema permitirá al usuario borrar todos los datos de un aspirante o estudiante cuando sea necesario, de esta manera se eliminan por completo los datos de la persona.
Identificación del requerimiento	RF10
Nombre del requerimiento	Mostrar datos de un aspirante o estudiante.
Descripción del requerimiento	El sistema permitirá al usuario visualizar todos los datos disponibles del perfil característico de una persona.

Identificación del requerimiento	RF11
Nombre del requerimiento	Calcular individualmente la posibilidad de deserción.
Descripción del requerimiento	El sistema permitirá al usuario calcular y mostrar la posibilidad de deserción de un individuo específico.
Identificación del requerimiento	RF12
Nombre del requerimiento	Importar datos de un grupo de personas.
Descripción del requerimiento	El sistema permitirá al usuario agregar a la base de datos de la plataforma varios perfiles de personas en una sola operación, los datos deben encontrarse en un archivo de hoja de cálculo con un formato que se describirá en el manual de la aplicación. La plataforma extrae los datos de la hoja de cálculo y registra los aspirantes o estudiantes encontrados en el archivo.
Identificación del requerimiento	RF13
Nombre del requerimiento	Buscar aspirantes o estudiantes.
Descripción del requerimiento	El sistema permitirá al usuario configurar los parámetros de búsqueda asociados a los atributos de un aspirante o estudiante para mostrar un número de perfiles de personas en los resultados de la consulta que cumplan con los criterios establecidos.
Identificación del requerimiento	RF14
Nombre del requerimiento	Mostrar resultados de búsqueda de aspirantes o estudiantes.
Descripción del requerimiento	El sistema permitirá al usuario visualizar un número de aspirantes o estudiantes que cumplen con unos filtros establecidos en una búsqueda. Se mostrará en el listado el número de identificación de la persona, nombres, apellidos, año de ingreso, periodo académico de ingreso y los datos de atributos establecidos en la búsqueda.
Identificación del requerimiento	RF15
Nombre del requerimiento	Borrar grupo de resultados

Descripción del requerimiento	El sistema permitirá al usuario eliminar un grupo de perfiles de personas pertenecientes a un resultado de búsqueda.
Identificación del requerimiento	RF16
Nombre del requerimiento	Clasificar un grupo de personas por posibilidad de deserción.
Descripción del requerimiento	El sistema permitirá al usuario clasificar un grupo de perfiles de personas pertenecientes a un resultado de búsqueda entre posibles desertores y no desertores. Se muestra una tabla con los resultados de la clasificación y una gráfica que resume los resultados en la interfaz de usuario web.
Identificación del requerimiento	RF17
Nombre del requerimiento	Calcular la posibilidad de deserción histórica.
Descripción del requerimiento	El sistema permitirá al usuario calcular la posibilidad de deserción de los aspirantes y/o estudiantes desde un año y un periodo de académico especificado hasta otro posterior. Se mostrará una tabla y una gráfica para facilitar la interpretación de los resultados.
Identificación del requerimiento	RF18
Nombre del requerimiento	Generar informe sobre deserción.
Descripción del requerimiento	El sistema permitirá al usuario generar un informe en formato de archivo PDF sobre la deserción en el programa de ingeniería de sistemas de la universidad de Cundinamarca en la extensión Facatativá. El informe incluye tablas y gráficos para exponer la situación.
Identificación del requerimiento	RF19
Nombre del requerimiento	Exportar datos de perfiles de personas a un archivo de hoja de cálculo.
Descripción del requerimiento	El sistema permitirá al usuario exportar a un archivo de hoja de cálculo los datos de un grupo de personas registradas en el sistema y pertenecientes a un resultado de búsqueda.
Identificación del requerimiento	RF20
Nombre del requerimiento	Entrenar el algoritmo de clasificación.

Descripción del requerimiento	El sistema permitirá al usuario entrenar el algoritmo de minería de datos de la plataforma para ajustarlo con los datos disponibles de perfiles de deserción.
-------------------------------	---

Nota: RF son las siglas de requerimiento funcional

Especificación del sistema

En esta sección se especifica el propósito y las características fundamentales del sistema que se desarrolla en la presente investigación.

Definición del sistema

Es un sistema para la recopilación de datos de perfiles característicos de estudiantes y la clasificación de los perfiles en cuanto a la posibilidad de deserción teniendo en cuenta la dimensión institucional, académica, socio-económica e individual de los estudiantes o aspirantes al programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, extensión Facatativá.

Tabla 8

Definición del sistema

P(ropietarios)	Universidad de Cundinamarca
A(ctores)	Docentes y administrativo encargados del proceso de selección de aspirantes y el análisis de la deserción estudiantil
T(ransformaciones)	Se transforman perfiles de estudiantes universitarios compuestos de 32 variables en una variable que indica si un estudiante es un posible desertor o no lo es.
C(lientes, beneficiarios o víctimas)	Docentes y administrativos encargados del proceso de selección de aspirantes. Estudiantes y aspirantes del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, extensión Facatativá.
R(estricciones o marco regulatorio del sistema)	La interfaz de usuario de la plataforma web estará diseñada para funcionar óptimamente en navegadores web de computadoras de escritorio y computadores portátiles.

	<p>Se usará el lenguaje de programación Python del lado de servidor y se usará el sistema gestor de base de datos Mysql.</p> <p>Se usarán los frameworks de desarrollo Flask y Flask-restless para Python. Se especificarán en archivos de texto las extensiones de Flask y los paquetes CONDA que sea necesario instalar.</p> <p>Se recomienda tener instalado en el equipo de destino de la plataforma el entorno Anaconda para instalar las librerías usadas en el proceso de minería de datos.</p> <p>Se recomienda un navegador web con soporte de HTML5, CSS3 y JavaScript para el acceso a la interfaz de usuario web.</p> <p>Se debe usar el protocolo HTTP para realizar peticiones a la interfaz REST.</p> <p>Es requerido el uso de JSON (JavaScript Object Notation) para manipular los datos en las peticiones de los recursos de la interfaz REST.</p> <p>Los servidores donde se instale el sistema deben ser capaces de atender consultas concurrentemente.</p>
W(eltanchauung, marco conceptual o declaración de intenciones del sistema)	<p>El sistema se denominará MIDUDEC, su intención es clasificar por posibilidad de deserción los perfiles de aspirantes o estudiantes del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca en la extensión de Facatativá para que se puedan tomar medidas que mitiguen el problema entre los afectados. El sistema almacenará los datos de los perfiles que se podrán ingresar y extraer usando una interfaz REST, archivos de hojas de cálculo y formularios HTML a través de un navegador web.</p>

Especificación del diseño

En este capítulo se encuentra la arquitectura de la plataforma web de deserción, la arquitectura está conformada por varios modelos UML 2.0 que se clasifican en tres grupos definidos como (Reinoso, 1999):

- **Dimensión Estática:** En este grupo se encuentra el modelo relacionado a la estructura del software.
- **Dimensión Dinámica:** En este grupo se encuentran los modelos relacionados a la interacción de objetos en el software.
- **Dimensión Funcional:** En este grupo se encuentran los modelos relacionados al comportamiento del software.

Modelo de entidad relación

El Modelo Entidad relación (MER) describe el diseño de la base de datos de la plataforma web de deserción realizada en MYSQL en donde se almacenará toda la información del software, la base de datos del aplicativo web consiste de 2 tablas identificadas cada una por un nombre y una llave primaria o Primary Key (PK), observar Figura 5.

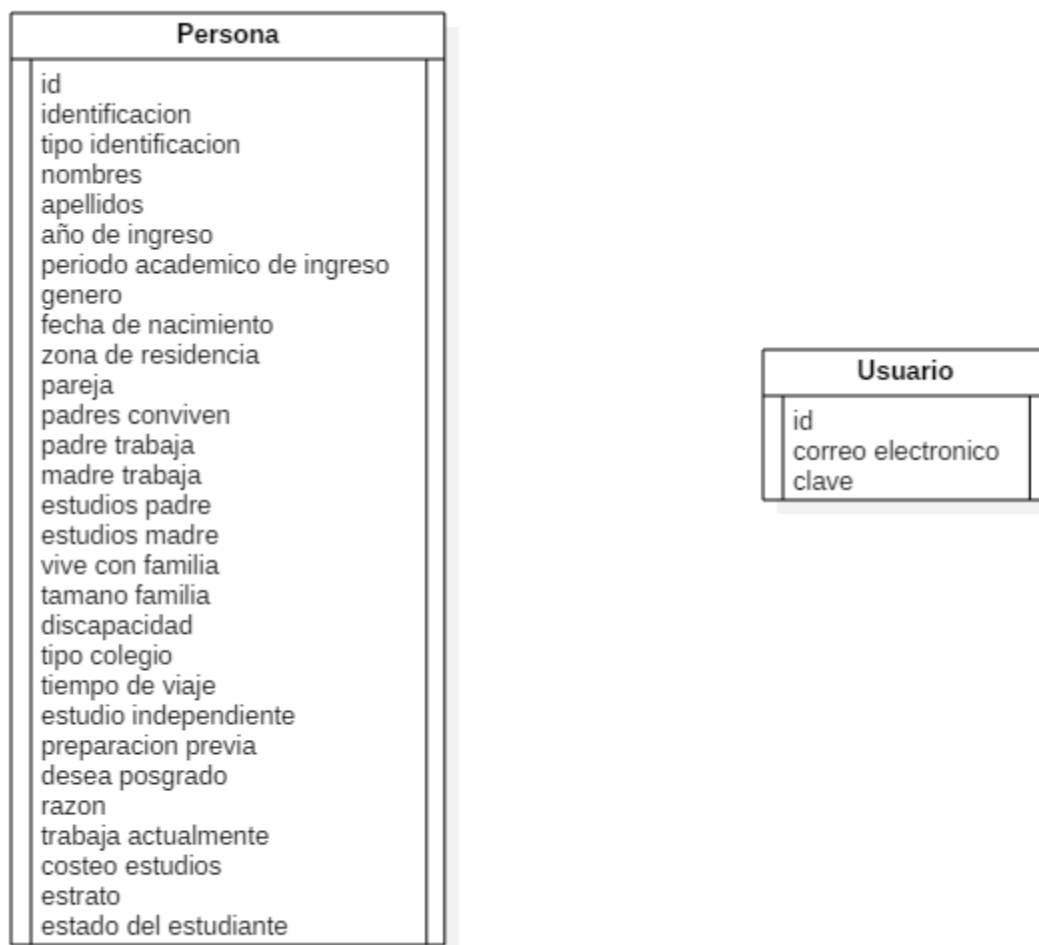


Figura 5. Modelo Entidad Relación

Descripción del modelo entidad relación

En la Tabla 9 se observará una breve descripción de cada una de las tablas existentes en el MER.

Tabla 9

Descripción del Modelo Entidad Relación

Tabla	Descripción
Persona	En esta tabla se almacenará la información de los perfiles de aspirantes o estudiantes de ingeniería de sistemas de la universidad de Cundinamarca.

Usuario	En esta tabla se almacena la información de acceso de las cuentas de usuario de la plataforma.
---------	--

Dimensión Estática

Diagrama de casos de uso

El Diagrama de Casos de uso contiene los actores que intervienen en el software y las funcionalidades que realiza cada uno de ellos.

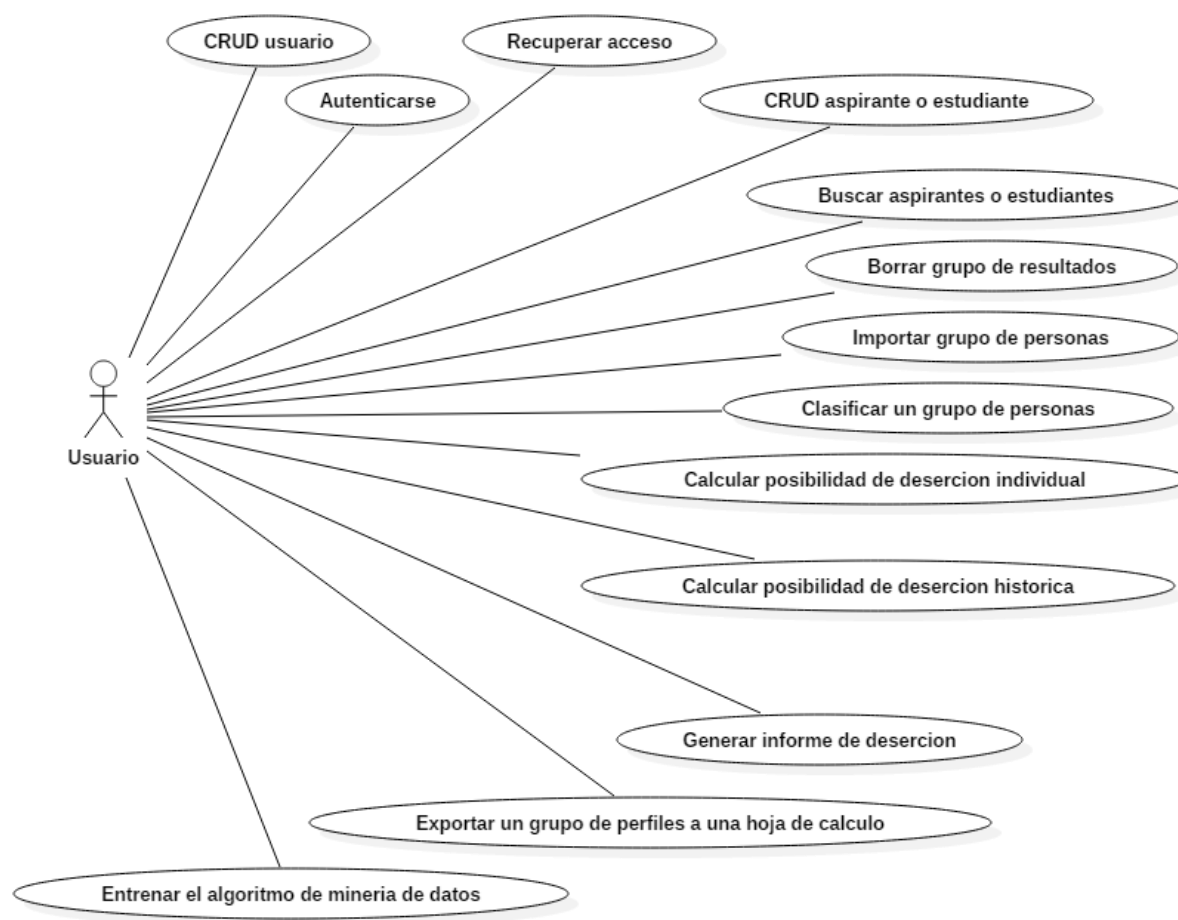


Figura 6. Diagrama de casos de uso

Descripción diagrama de casos de uso

En la Tabla 10 se describen las funcionalidades mostradas en el diagrama de casos de uso.

Tabla 10
Descripción del diagrama de casos de uso

Caso de uso	Descripción
CRUD usuario	Operaciones de lectura, registro, actualización y borrado de usuarios de la plataforma.
Autenticarse	Validar el acceso a los recursos de la aplicación con el correo electrónico y la contraseña del usuario.
CRUD aspirante o estudiante	Operaciones de lectura, registro, actualización y borrado de perfiles característicos de aspirantes o estudiantes.
Buscar aspirantes o estudiantes	Filtrar un grupo de aspirantes o estudiantes que cumplan unos valores dados para las diferentes características del perfil.
Borrar grupo de resultados	Borrar un grupo de aspirantes o estudiantes que pertenecen a una búsqueda.
Importar grupo de personas	Registrar un grupo de perfiles de personas almacenados archivo de hoja de cálculo.
Clasificar grupo de personas	Se calcula la posibilidad de deserción para un grupo de personas que conforman un resultado de búsqueda.
Calcular posibilidad de deserción individual	Se calcula la posibilidad de deserción de un aspirante o estudiante seleccionado de los resultados de búsqueda.
Calcular posibilidad de deserción histórica	Se calcula la posibilidad de deserción de un grupo de personas que conforman un resultado de búsqueda para un periodo de tiempo establecido.
Generar informe de deserción	Se calcula la posibilidad de deserción de todos los perfiles característicos y se genera un informe con los resultados en un archivo de formato PDF.
Exportar un grupo de perfiles a una hoja de calculo	Exportar los perfiles característicos de un resultado de búsqueda a un archivo de hoja de cálculo que puede ser descargado para su uso en otras actividades de la universidad.
Entrenar algoritmo de minería de datos	Ajustar a los datos de entrenamiento disponibles en la base de datos los valores de las fórmulas matemáticas del clasificador que se use como algoritmo de minería de datos para clasificar perfiles por posibilidad de deserción.
Recuperar acceso	Enviar los datos de acceso al correo electrónico del usuario respectivo.

Dimensión Dinámica

Diagrama de secuencia

El diagrama de secuencia muestra la interacción de objetos realizada en el sistema y explica detalladamente el modelo de casos de uso en el orden en que se aplican las funcionalidades.

Diagramas de secuencia del sistema MIDUDEC

Se realiza un diagrama de secuencia para cada caso de uso.

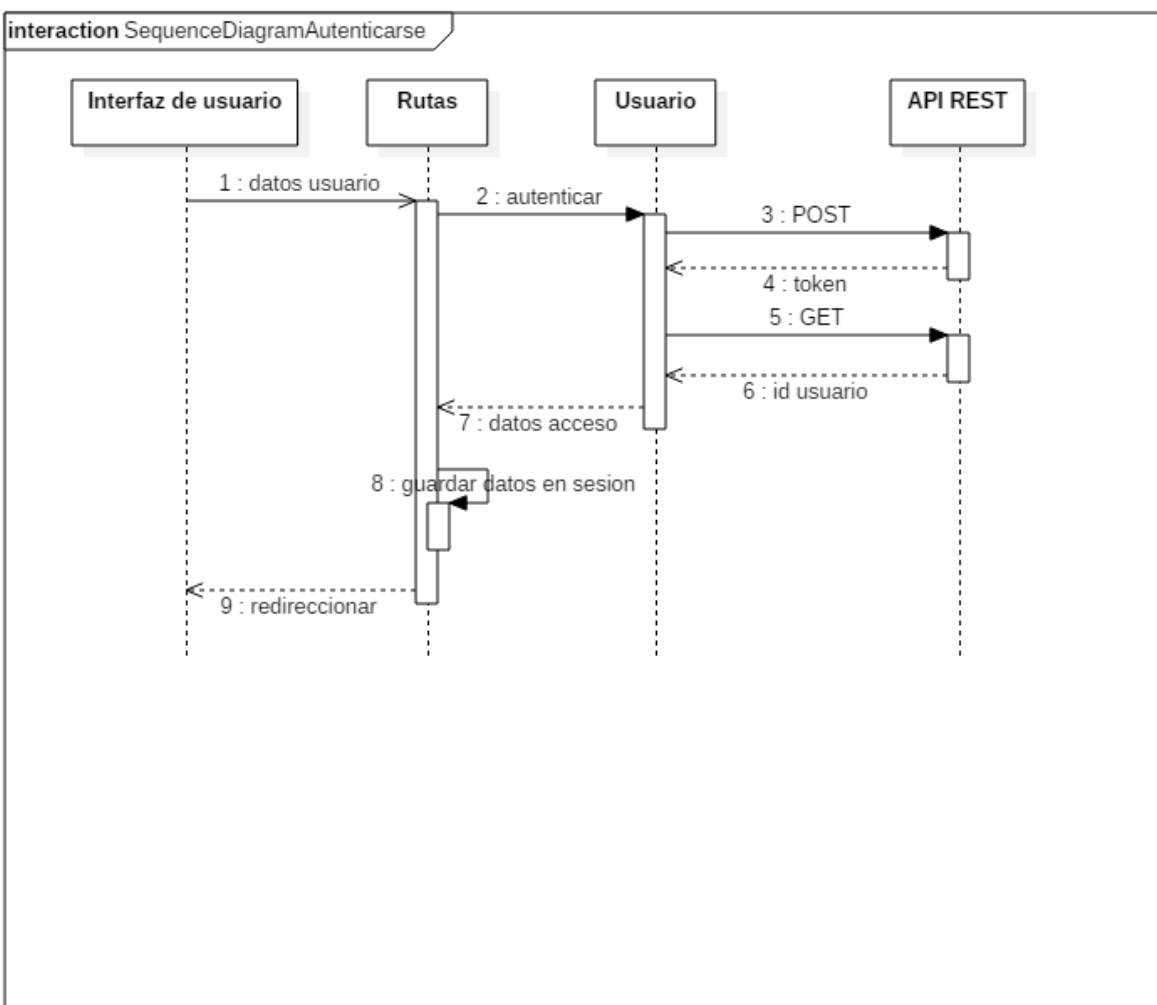


Figura 7. Diagrama de secuencia Autenticarse

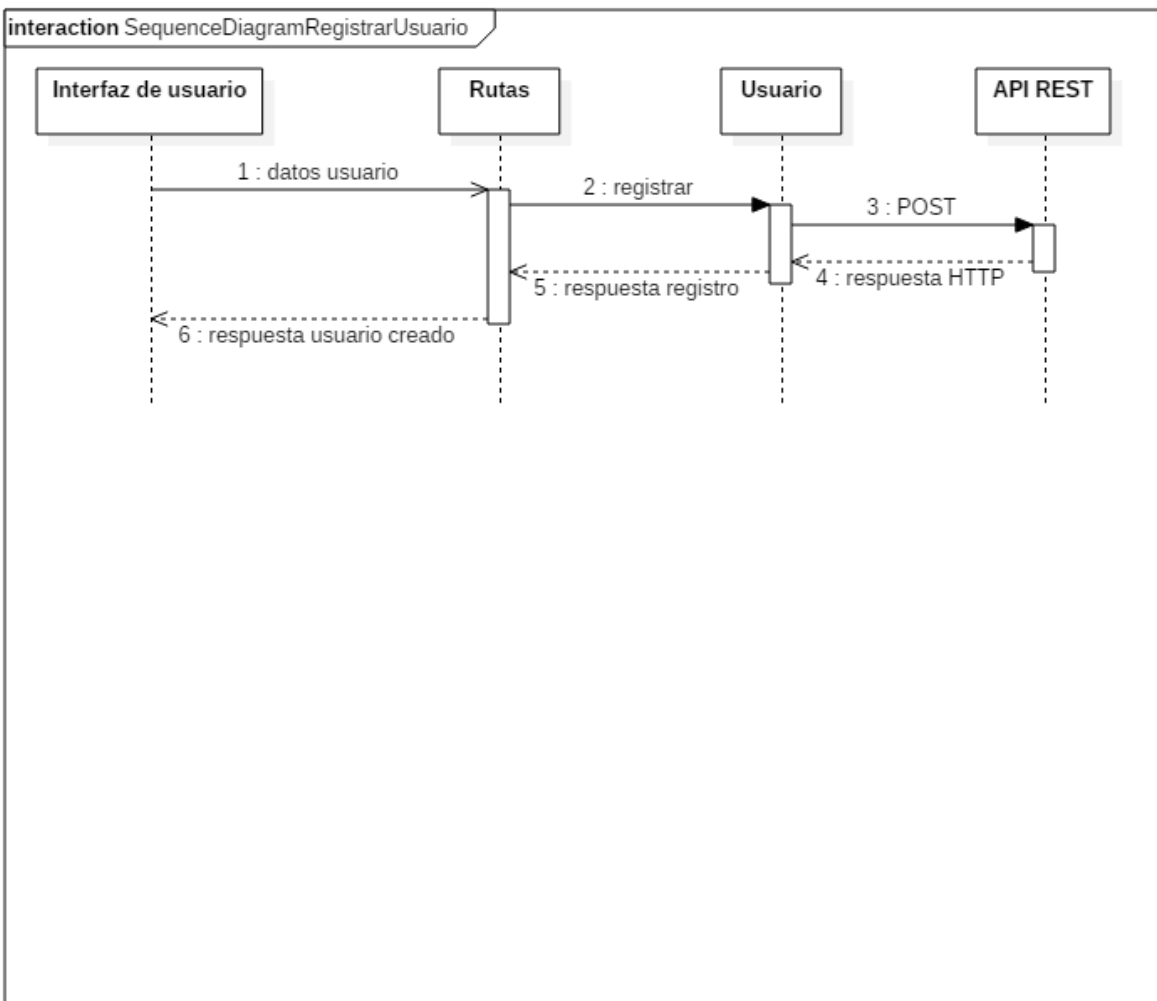


Figura 8. Diagrama de secuencia Registrar Usuario

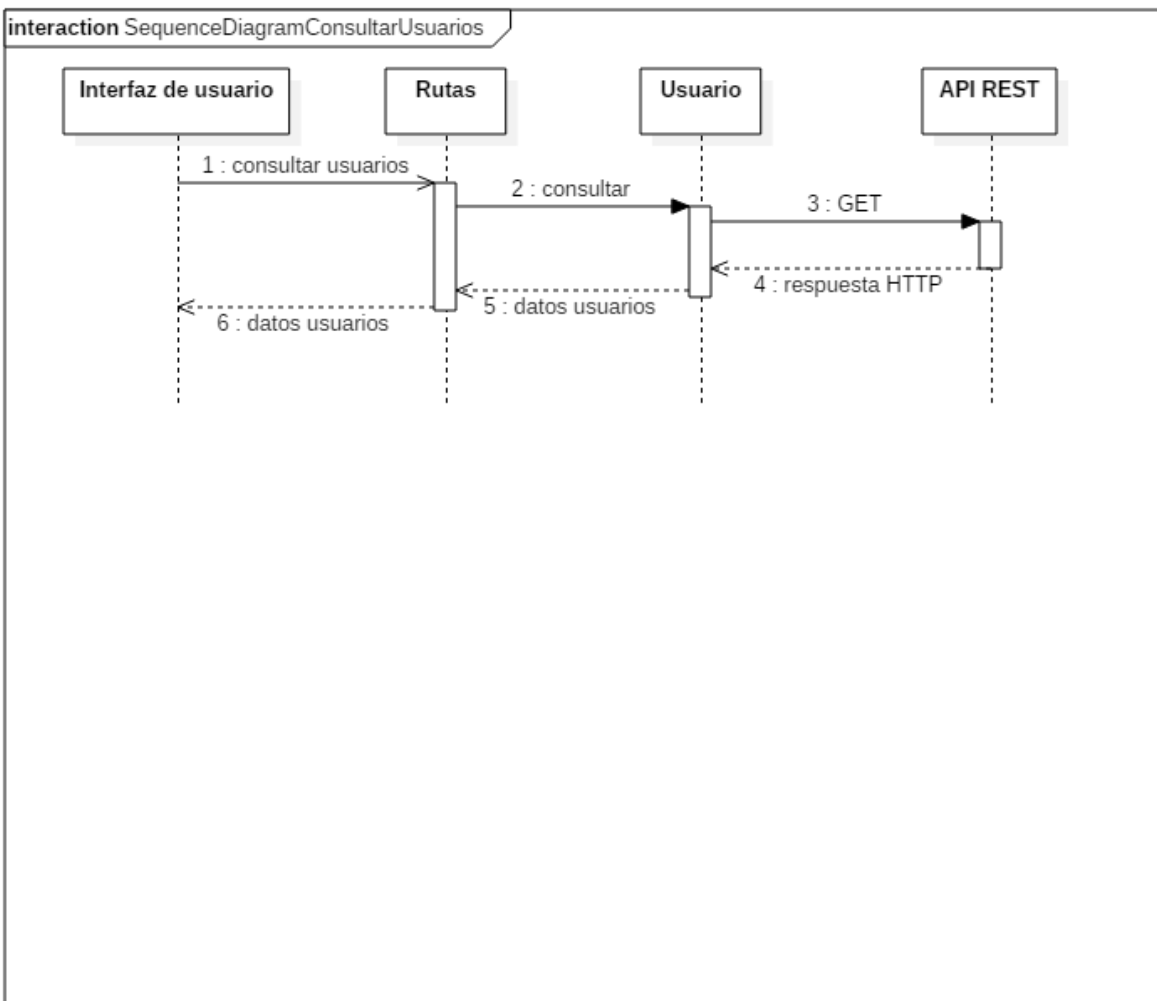


Figura 9. Diagrama de secuencia Consultar Usuarios

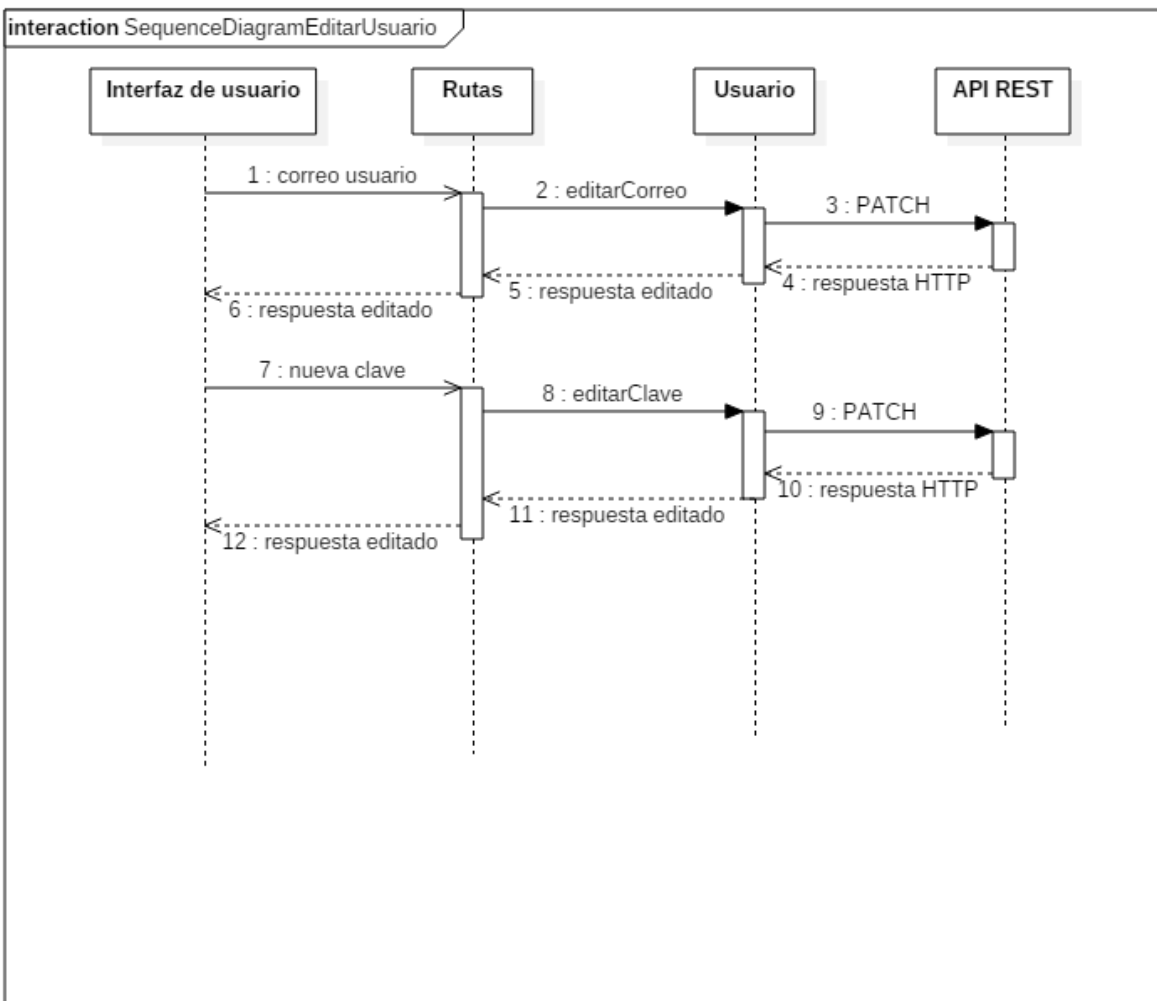


Figura 10. Diagrama de secuencia Editar Usuario

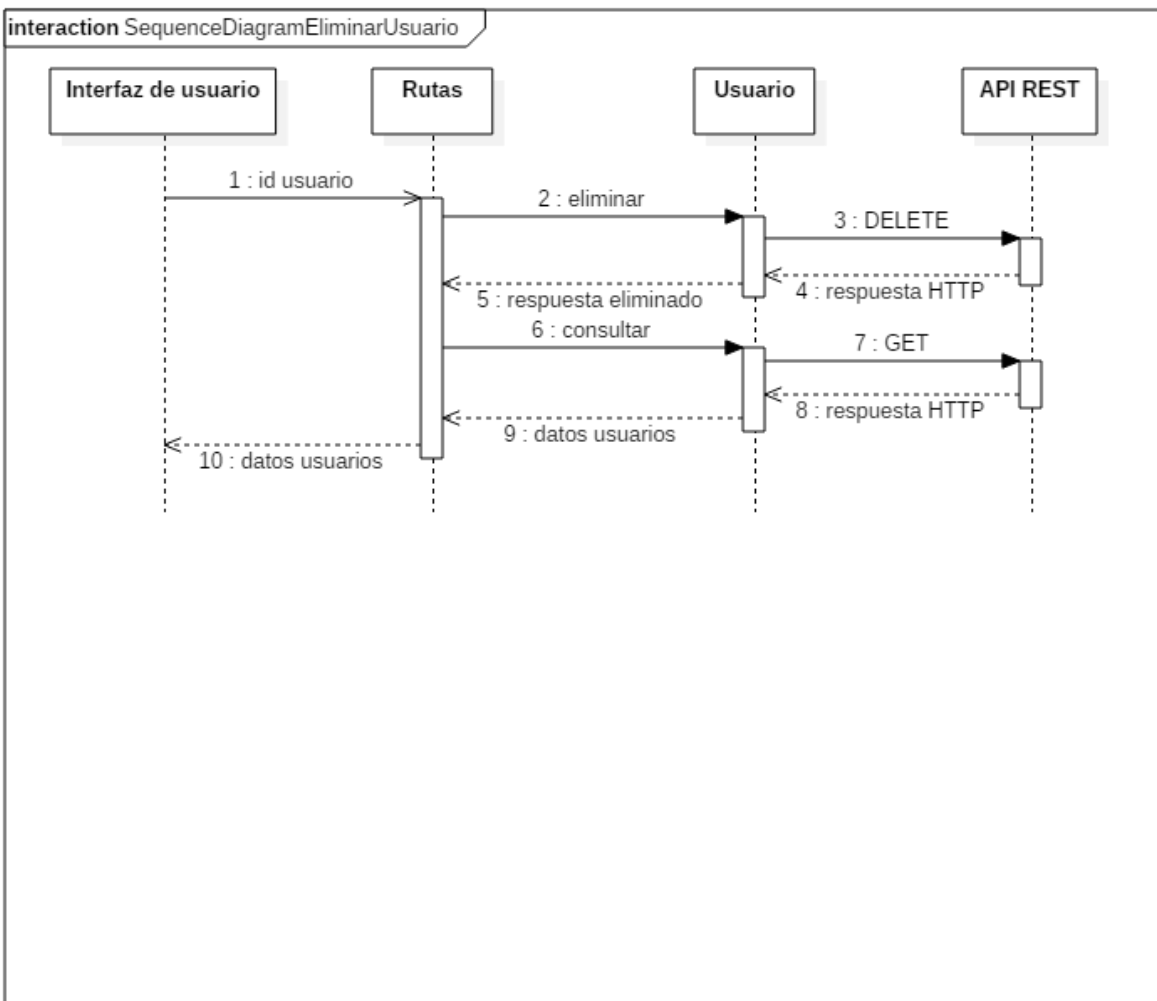


Figura 11. Diagrama de secuencia Eliminar Usuario

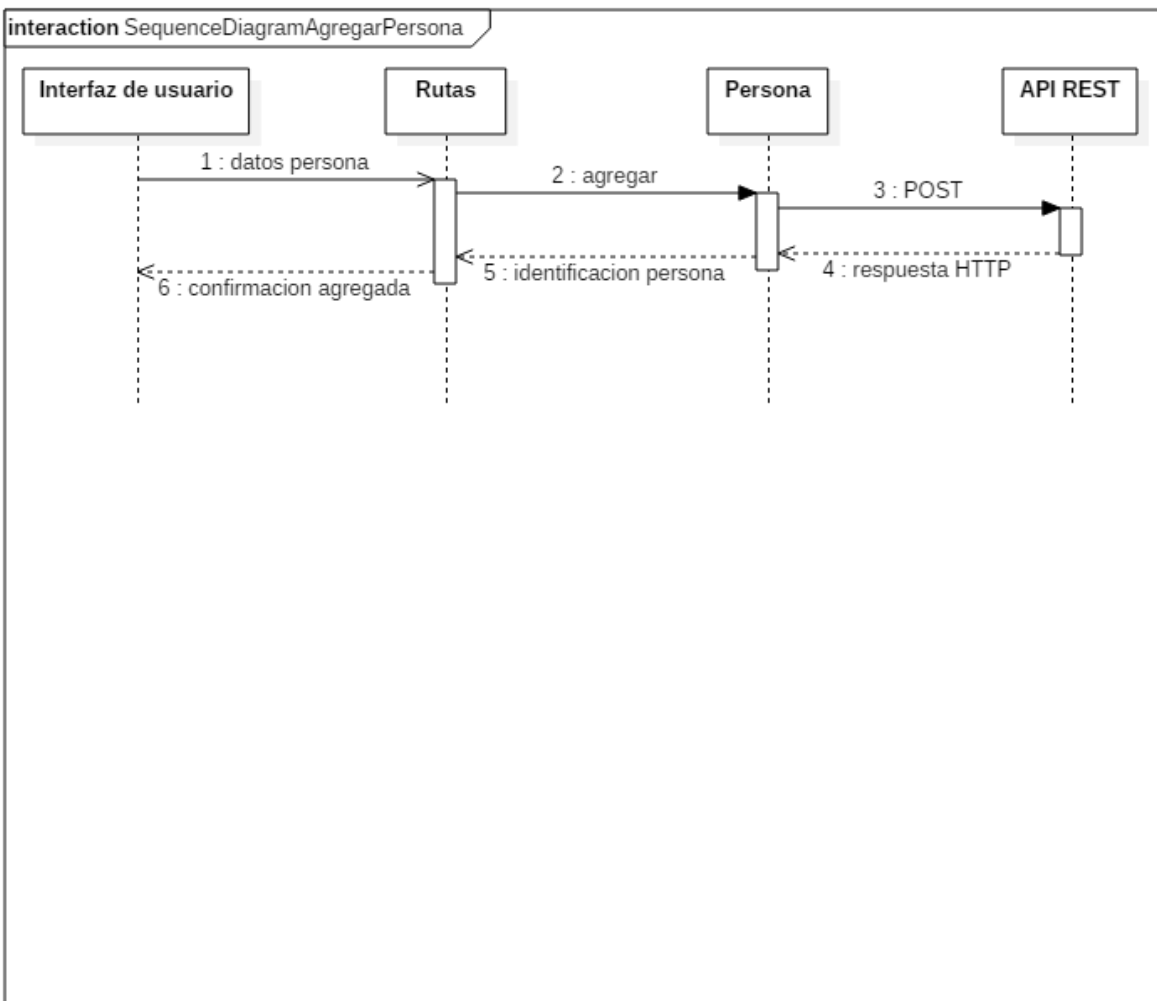


Figura 12. Diagrama de secuencia Agregar Persona

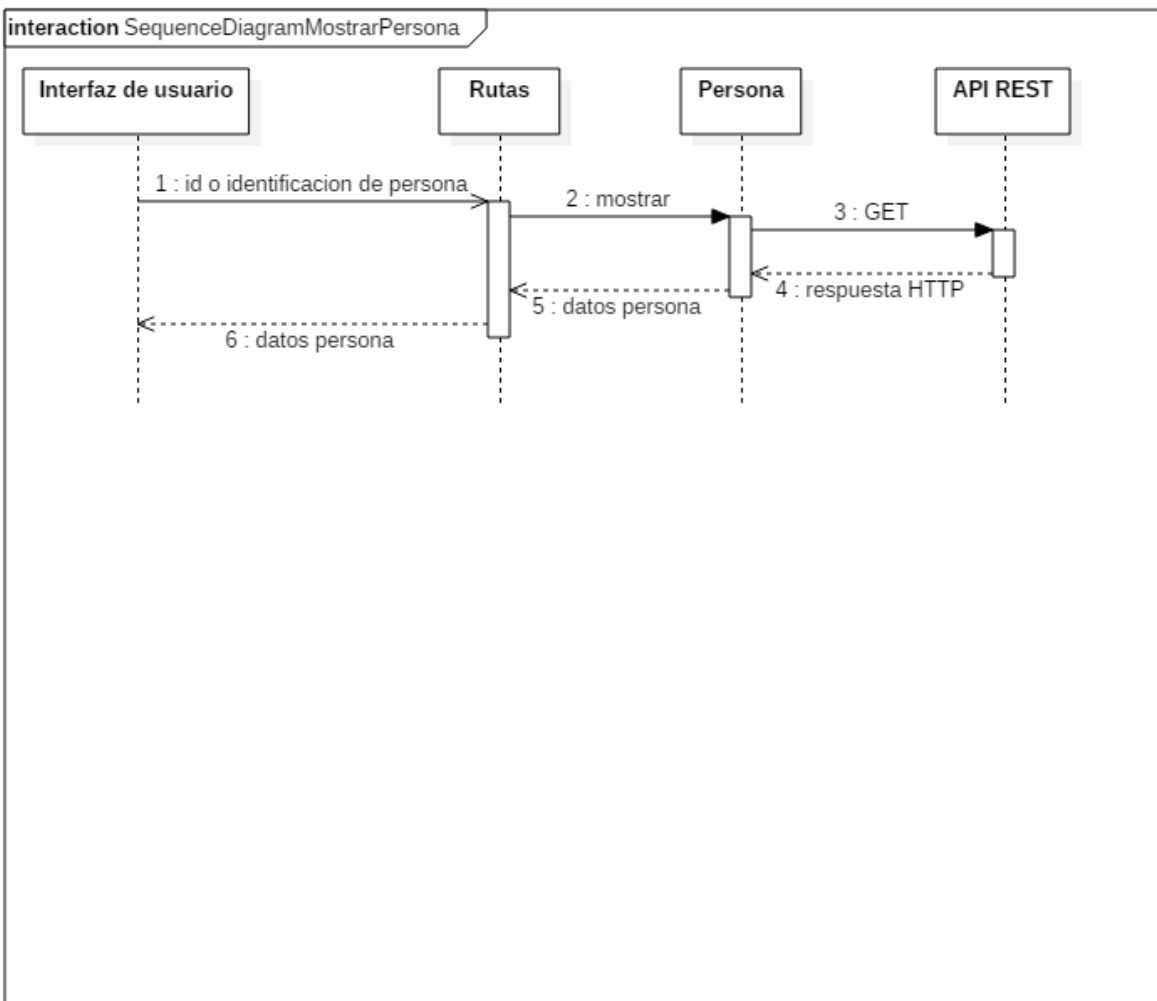


Figura 13. Diagrama de secuencia Mostrar Persona

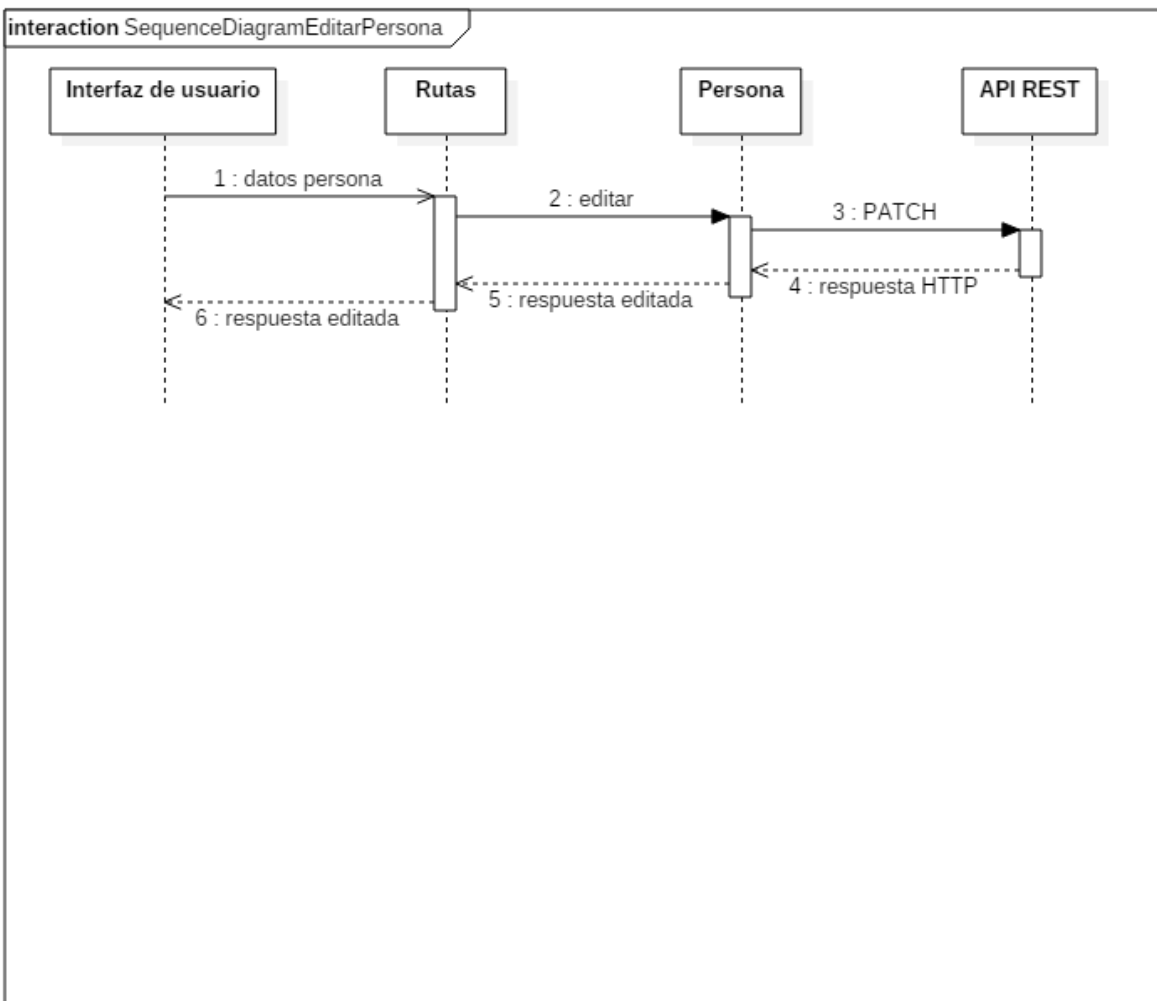


Figura 14. Diagrama de secuencia Editar Persona

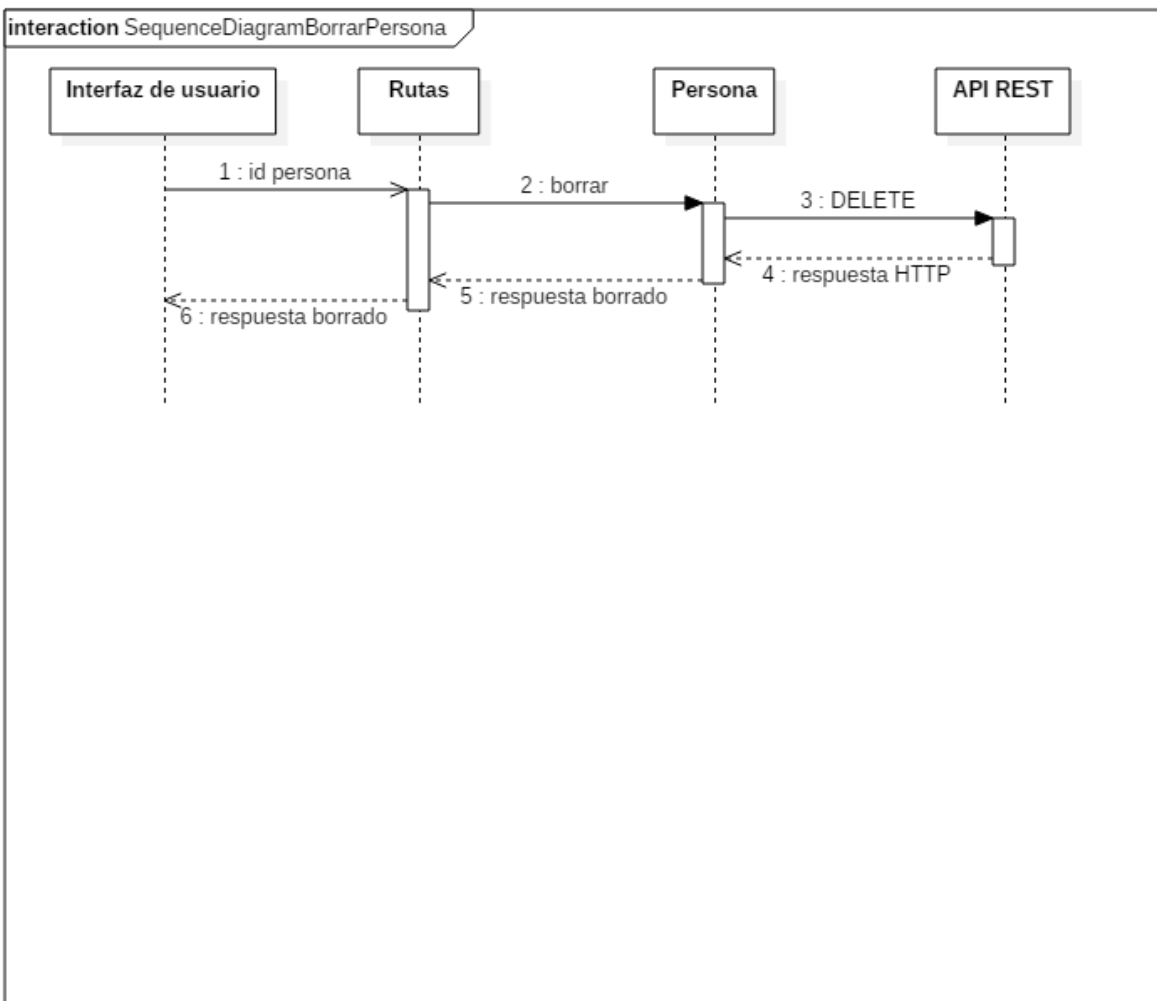


Figura 15. Diagrama de secuencia Borrar Persona

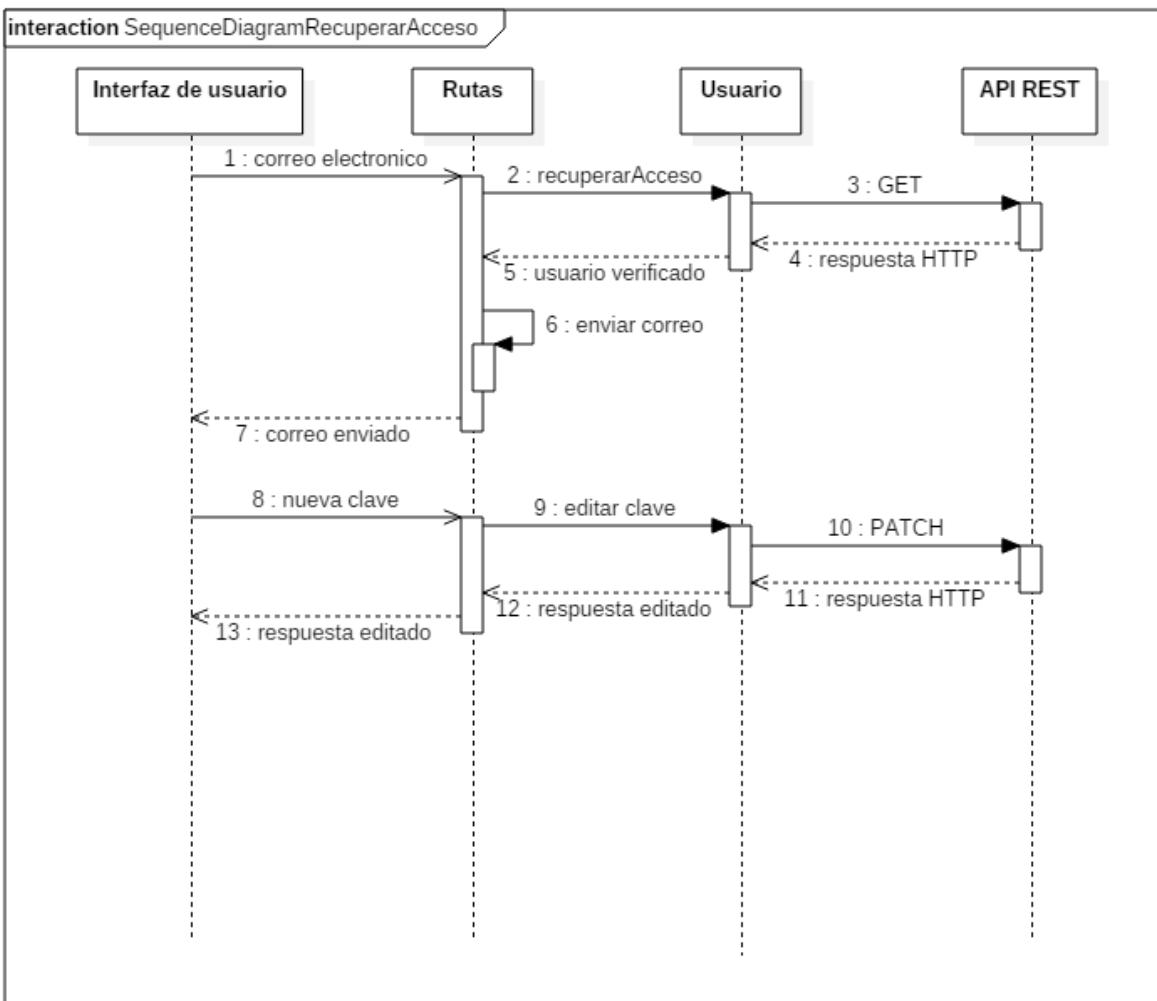


Figura 16. Diagrama de secuencia Recuperar Acceso

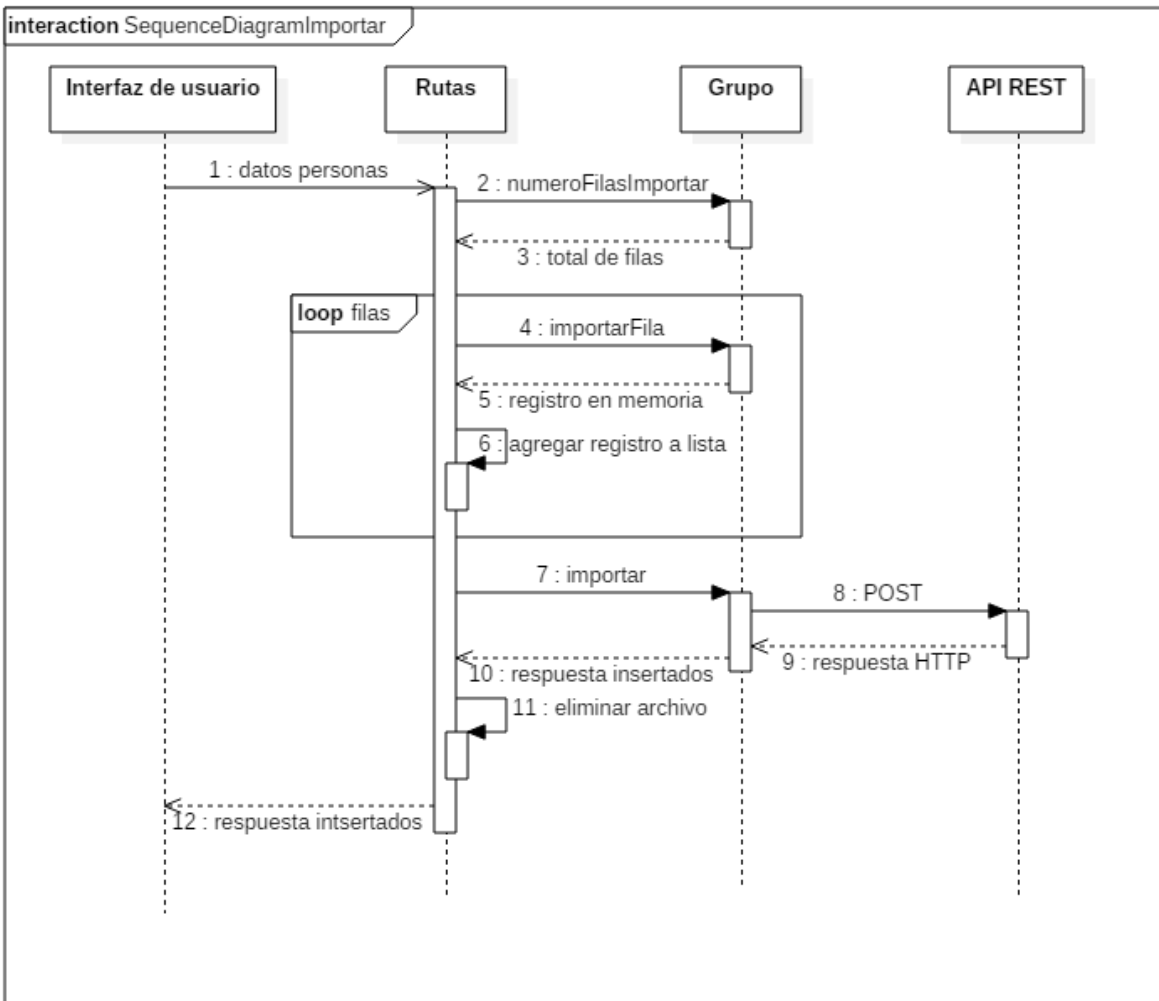


Figura 17. Diagrama de secuencia Importar

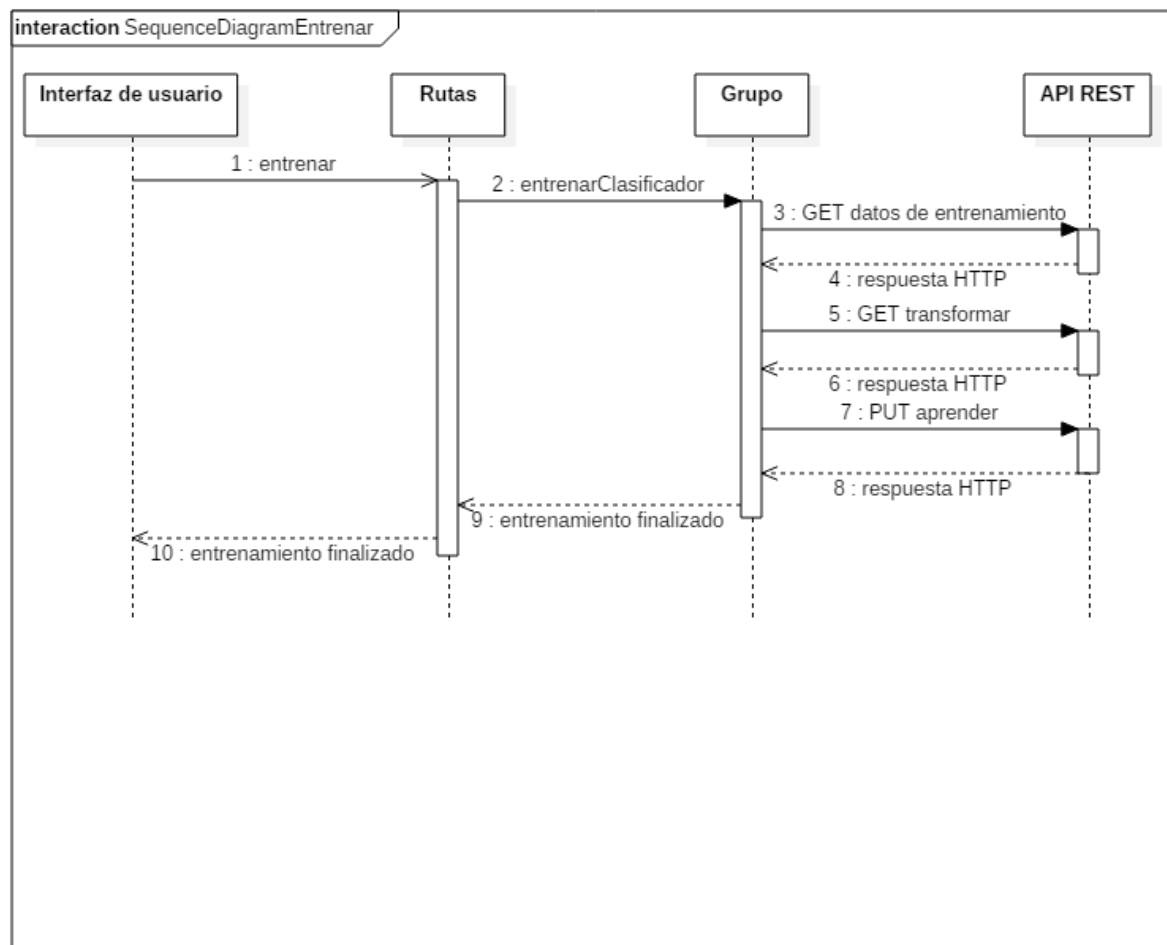


Figura 18. Diagrama de secuencia Entrenar

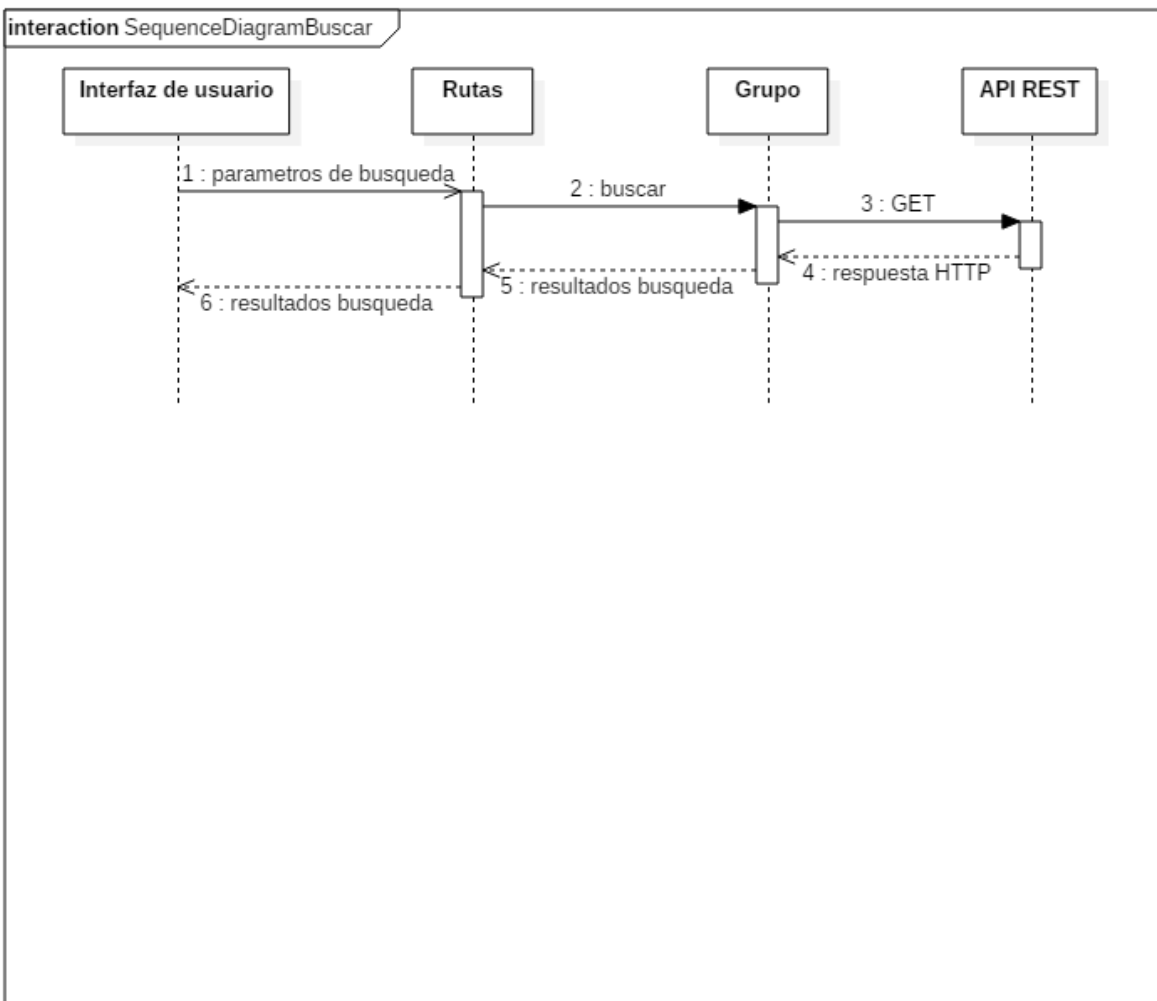


Figura 19. Diagrama de secuencia Buscar

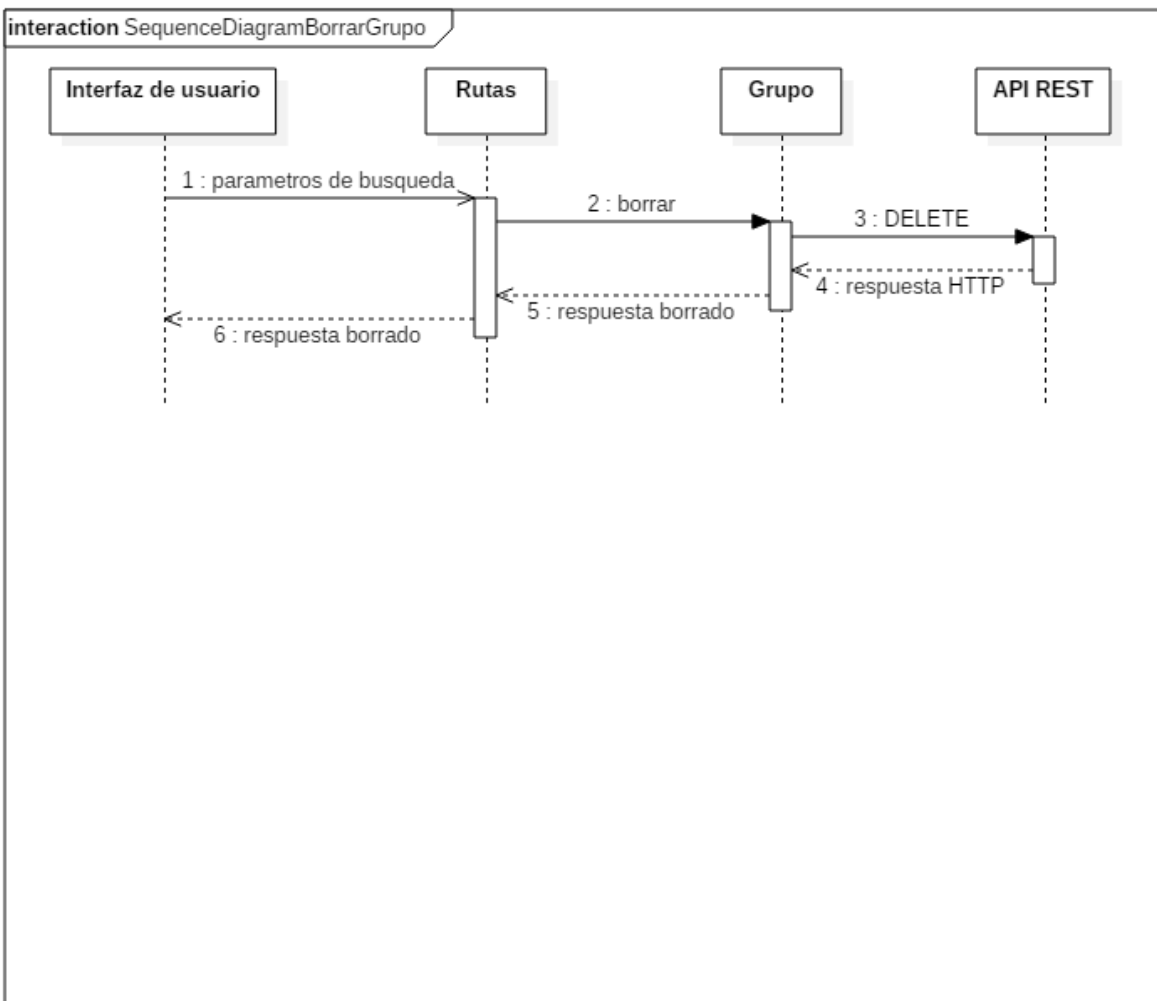


Figura 20. Diagrama de secuencia Borrar Grupo

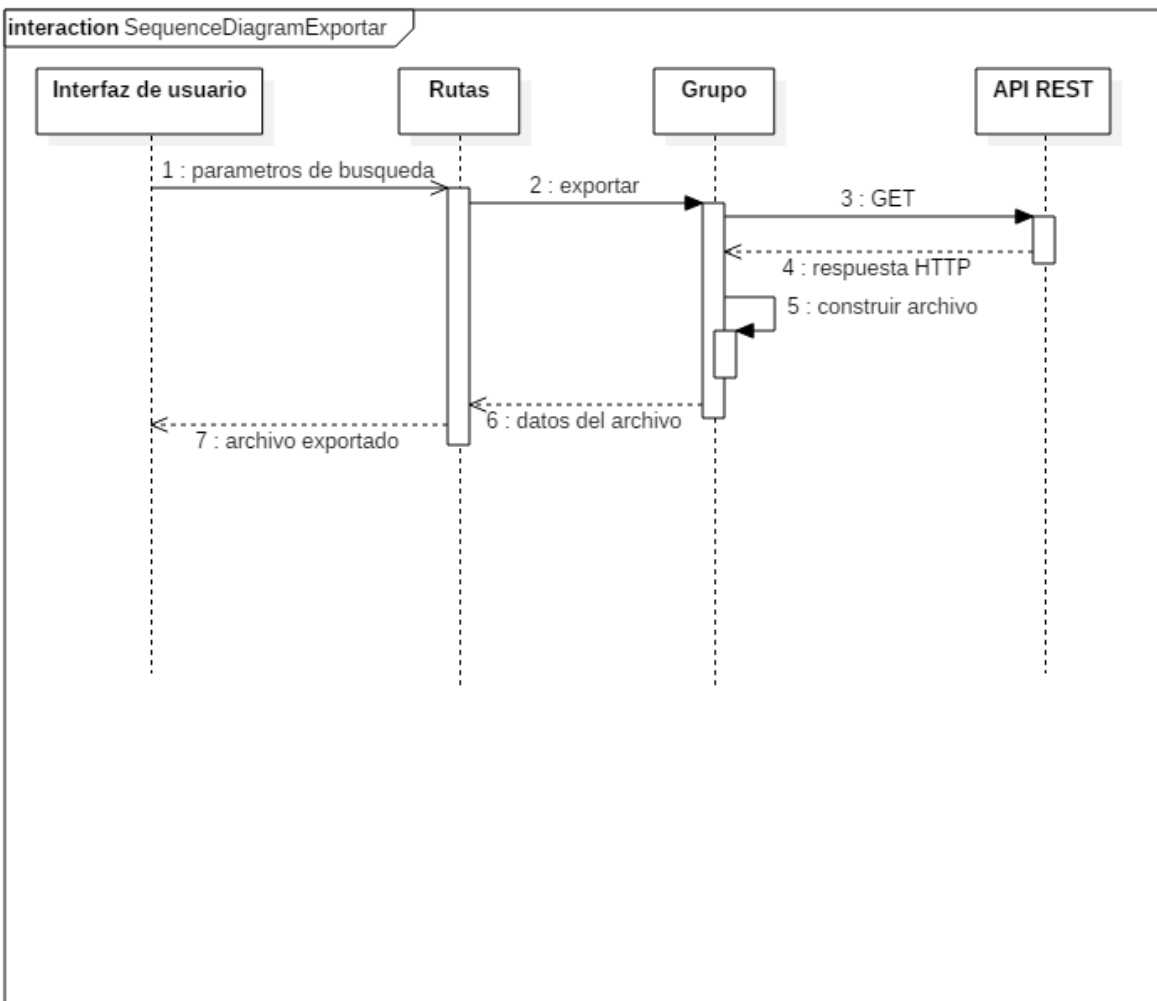


Figura 21. Diagrama de secuencia Exportar

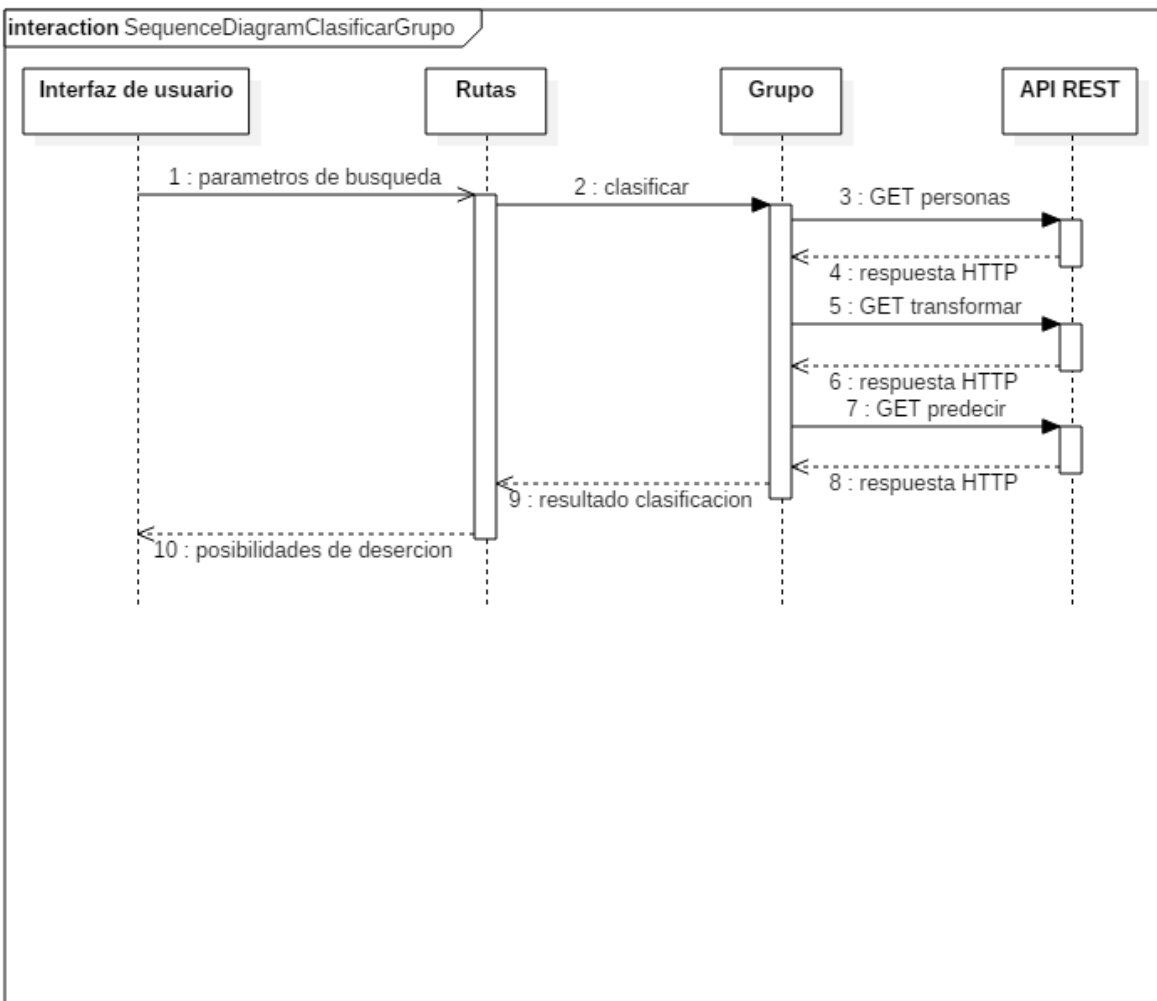


Figura 22. Diagrama de secuencia Clasificar Grupo

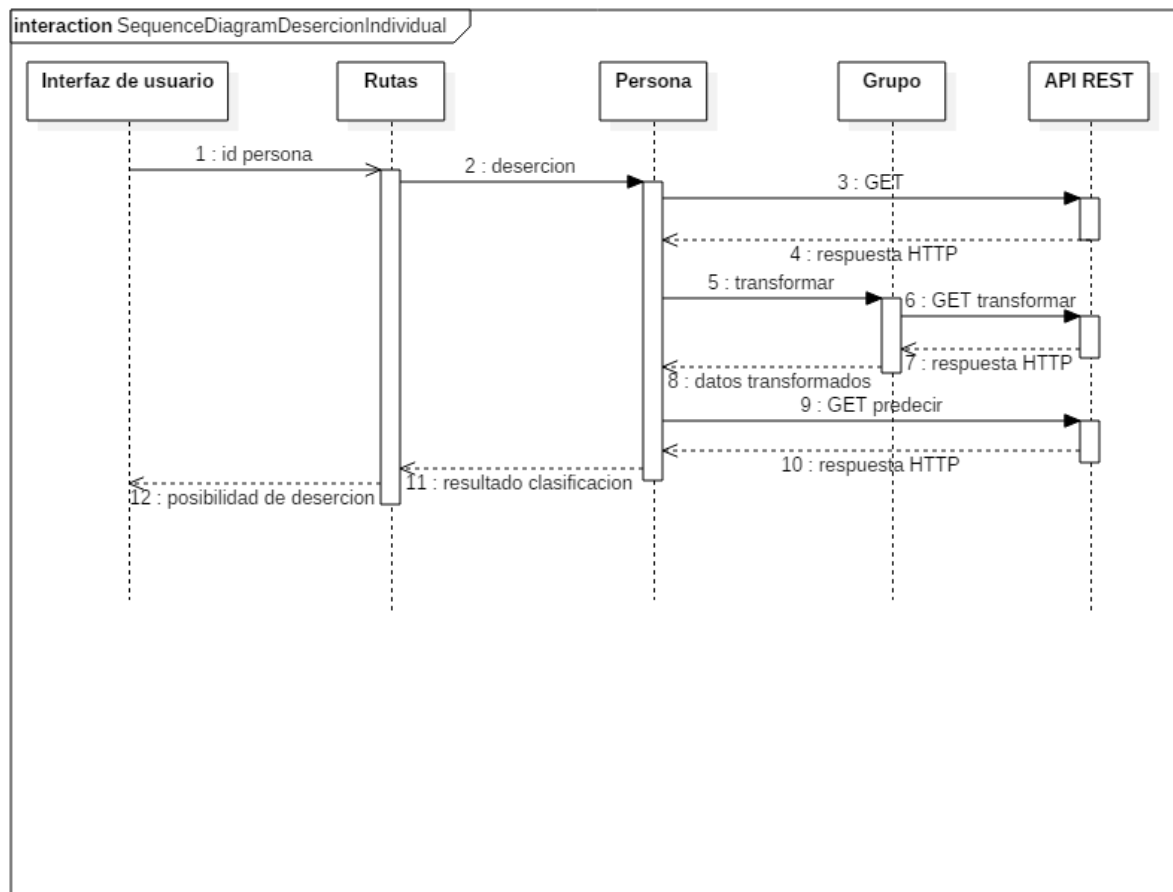


Figura 23. Diagrama de secuencia Deserción Individual

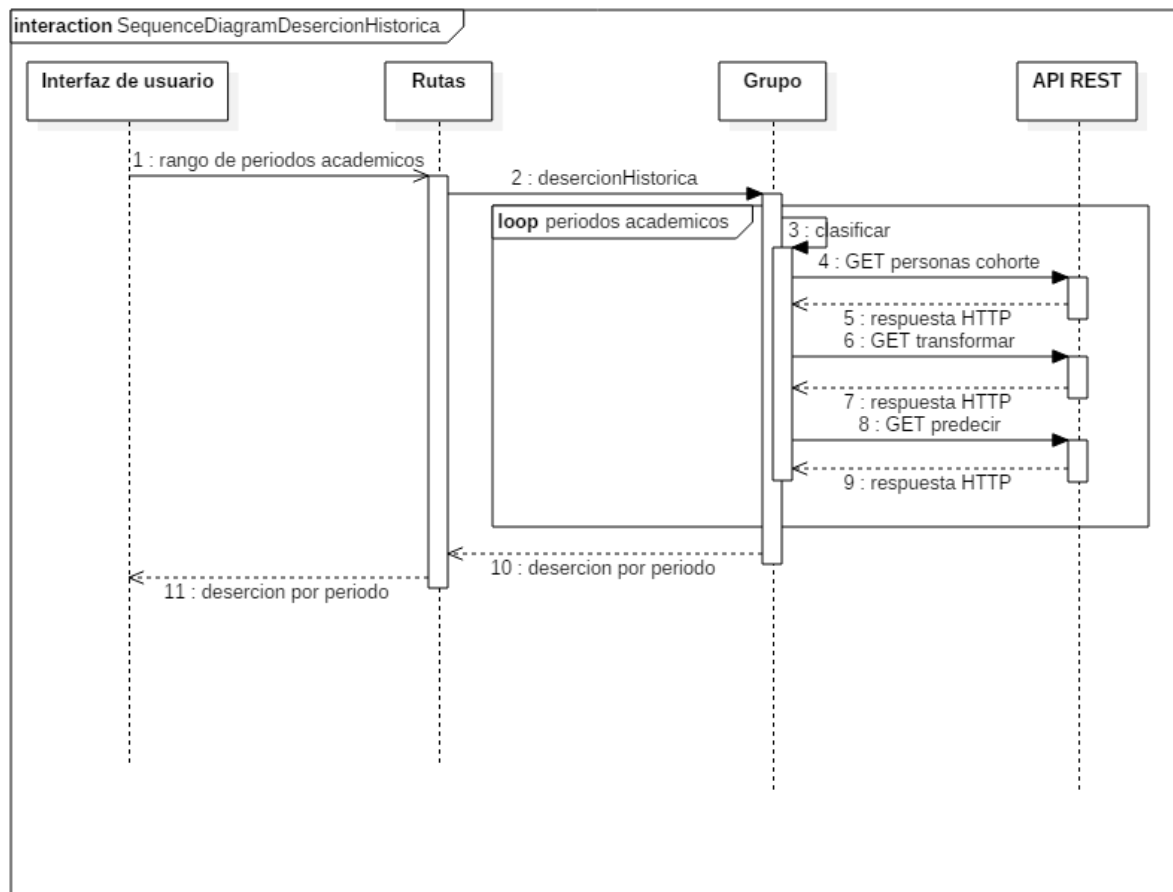


Figura 24. Diagrama de secuencia Deserción Histórica

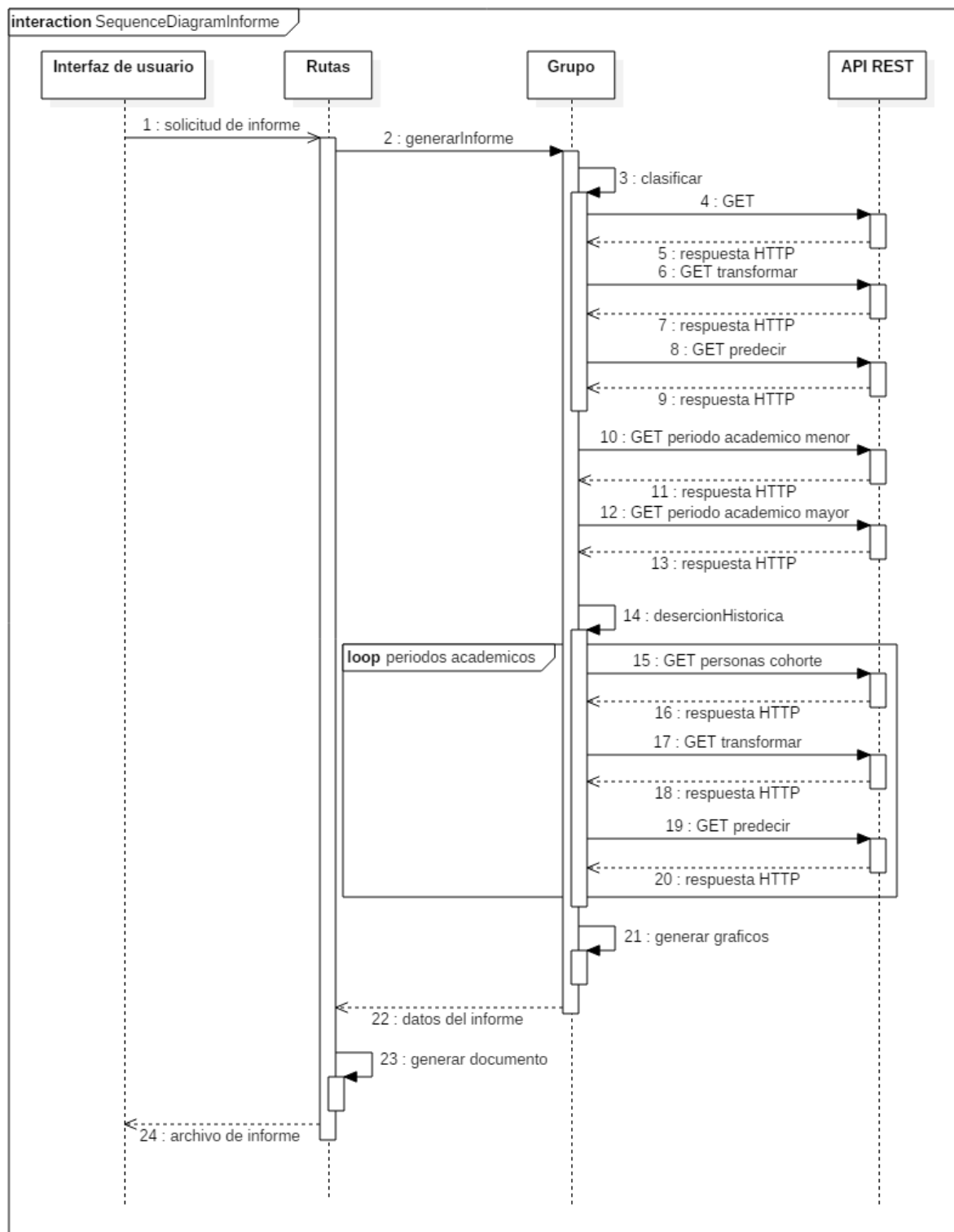


Figura 25. Diagrama de secuencia Informe

Descripción de los diagramas de secuencia

En la Tabla 11 se describen los objetos de los diagramas de secuencia.

Tabla 11
Descripción de los diagramas de secuencia

Objeto	Descripción
Interfaz de usuario	La interfaz de usuario web hace peticiones HTTP al sistema de rutas del servidor web.
Rutas	El sistema de rutas es un módulo del servidor web que hace llamadas a objetos del controlador dependiendo de la URL solicitada por el navegador web.
Usuario	Usuario es una clase del controlador que encapsula los atributos y métodos de un usuario.
Persona	Persona es una clase del controlador que encapsula los atributos y métodos de un aspirante o estudiante.
Grupo	Grupo es una clase del controlador que encapsula los atributos y métodos de un grupo mediante la composición de varios objetos de tipo Persona.
API REST	El API REST es un módulo que recibe peticiones HTTP, en este caso de los controladores del servidor web y realiza la consulta requerida en la base de datos para retornar los datos solicitados. La conexión con la base de datos se realiza por medio del ORM (object-relational mapper) SQLAlchemy.

Dimensión Funcional

Diagrama de actividades

El diagrama de actividades hace parte de la dimensión funcional que se encarga de explicar todo el comportamiento de los usuarios en su interacción con el software. Este diagrama muestra el flujo de información con el que usuario tendrá acceso a la aplicación. A continuación, se exponen los diagramas de actividades para cada caso de uso.

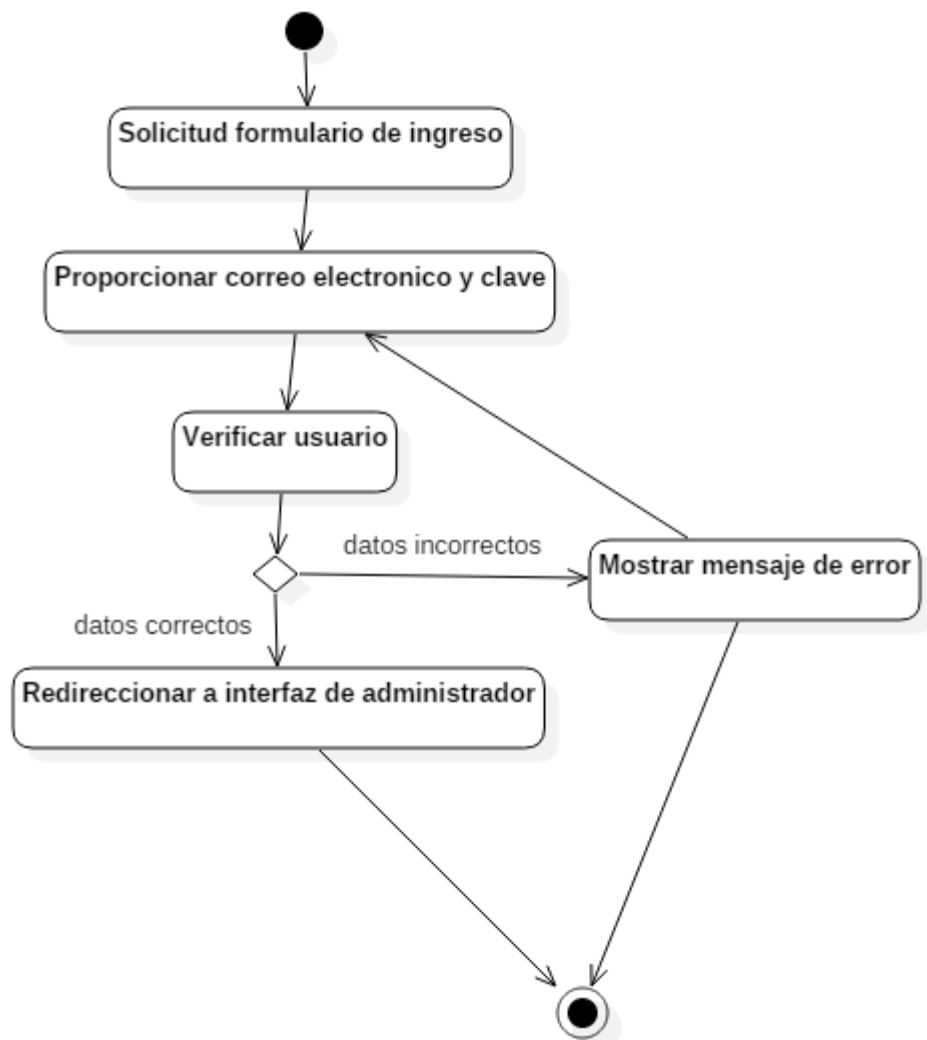


Figura 26. Diagrama de actividad Autenticarse

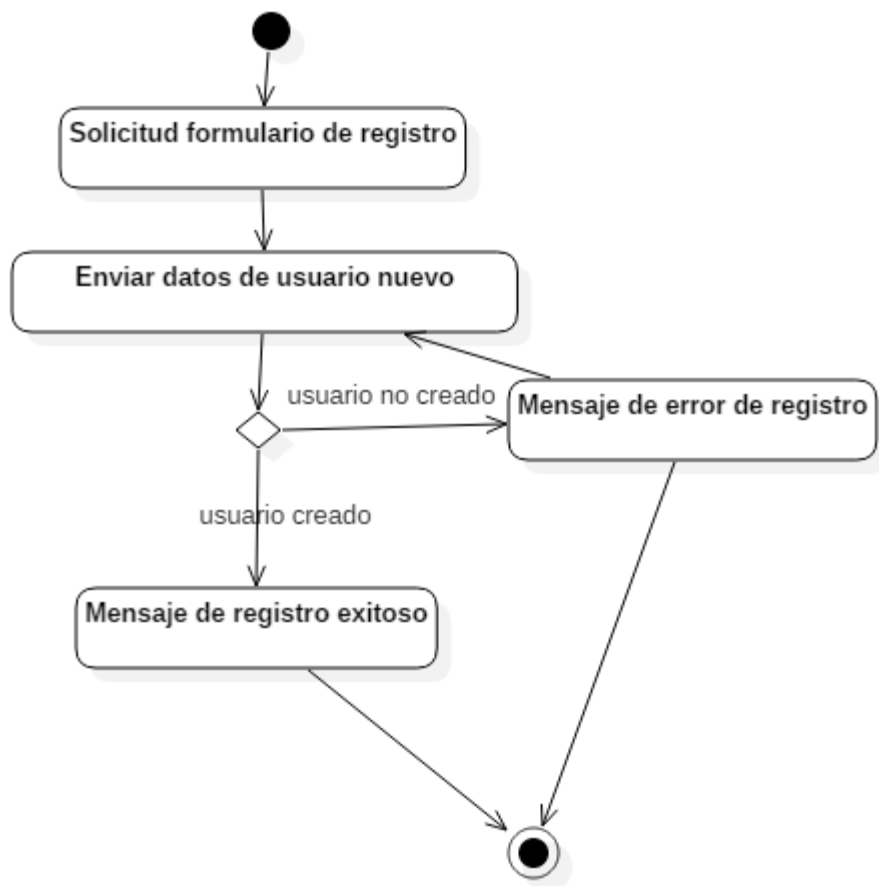


Figura 27. Diagrama de actividad Registrar Usuario

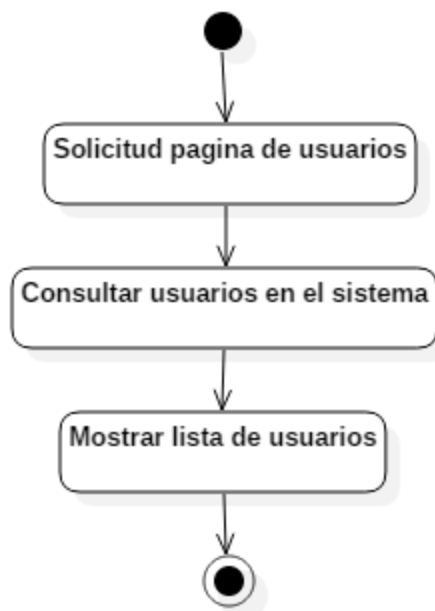


Figura 28. Diagrama de actividad Consultar Usuarios

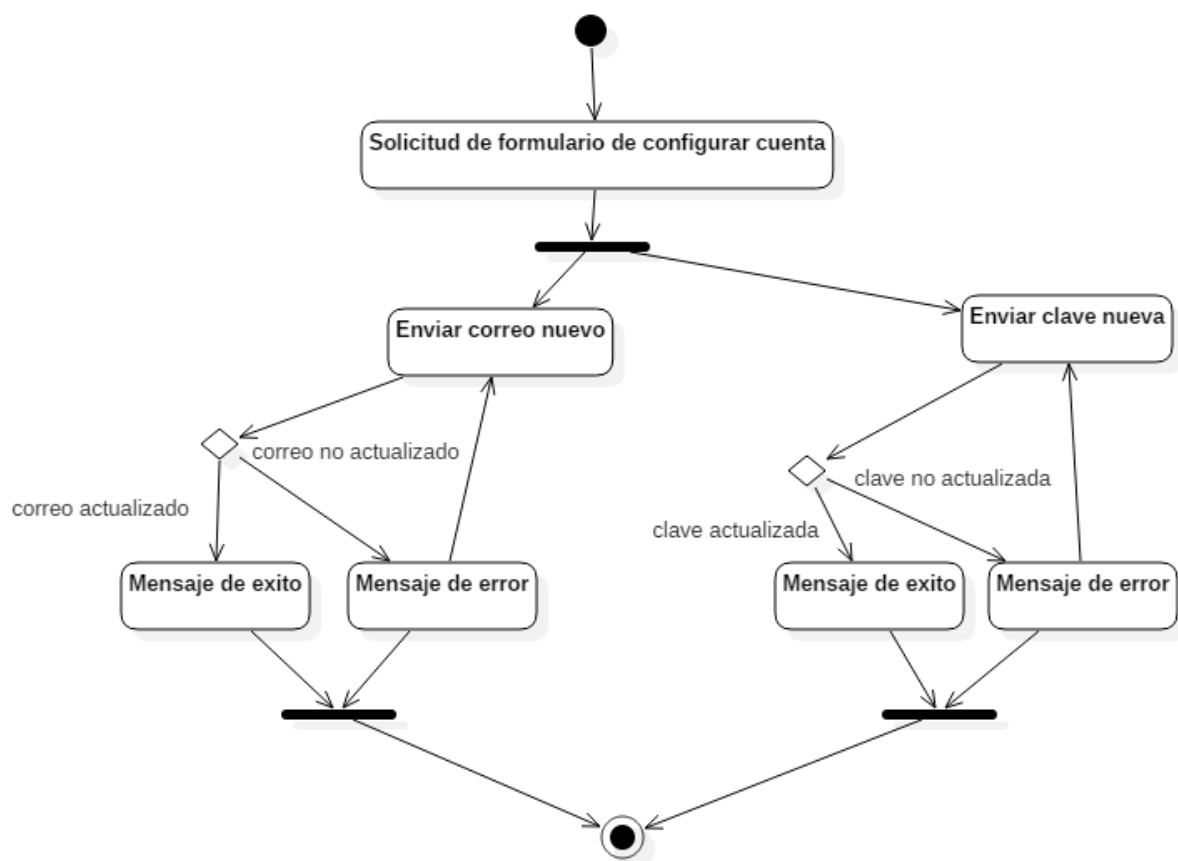


Figura 29. Diagrama de actividad Editar Usuario

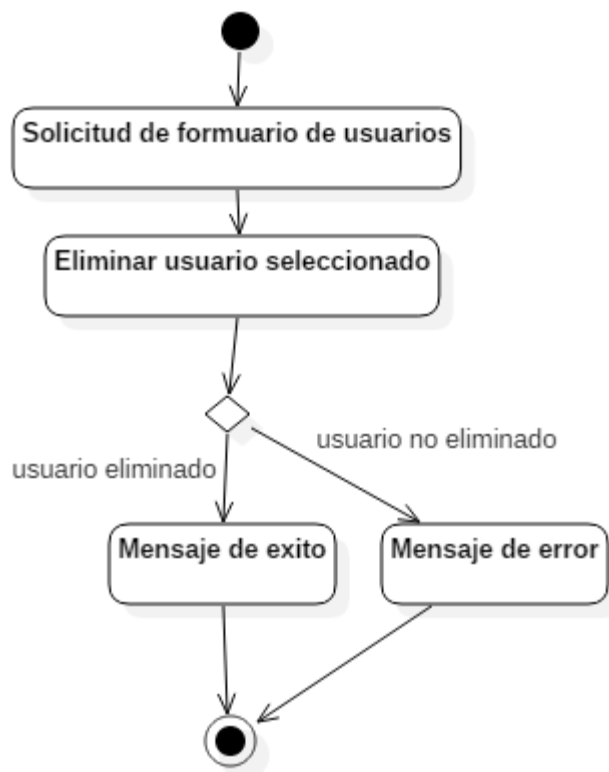


Figura 30. Diagrama de actividad Eliminar Usuario

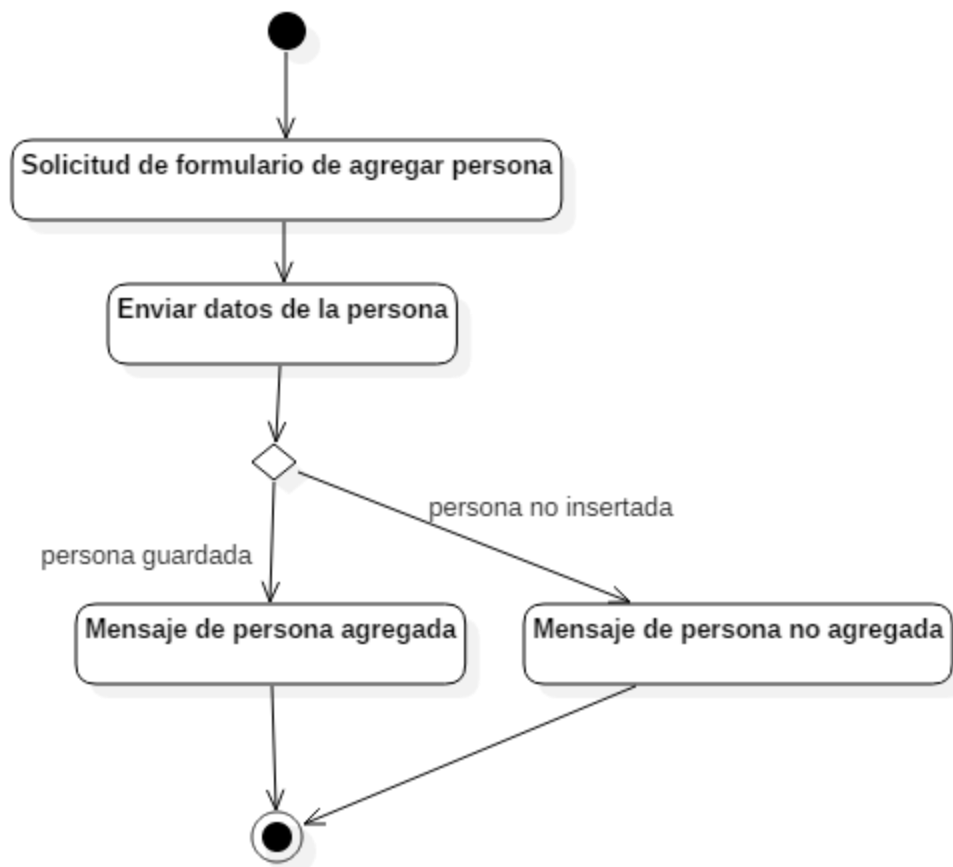


Figura 31. Diagrama de actividad Agregar Persona

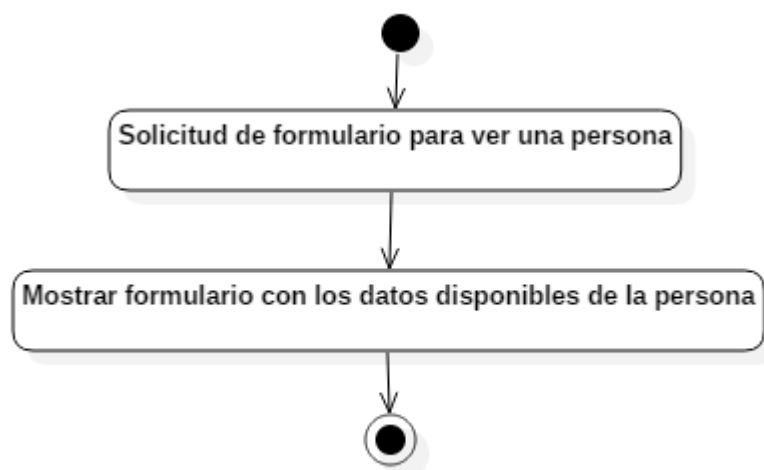


Figura 32. Diagrama de actividad Mostrar Persona

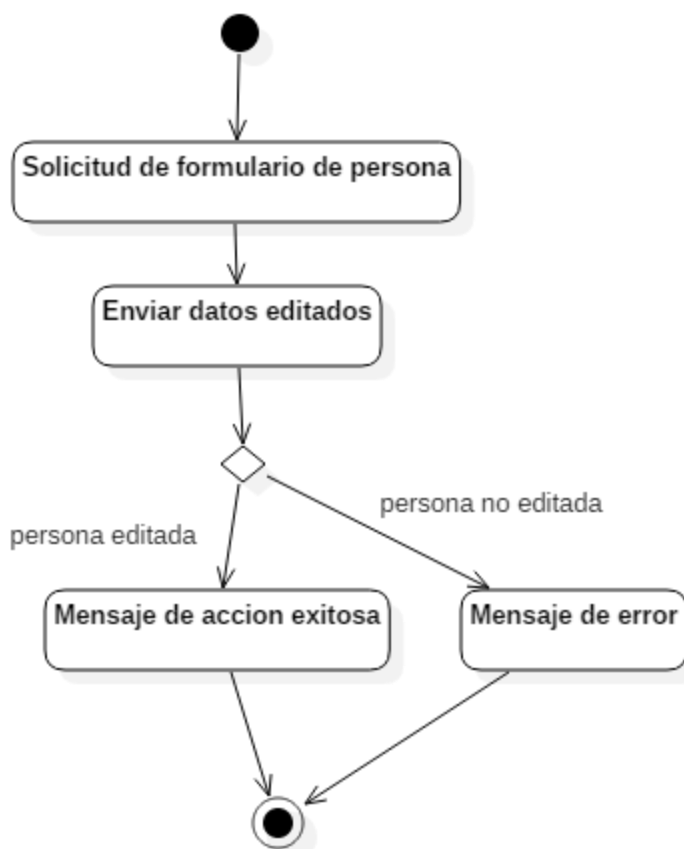


Figura 33. Diagrama de actividad Editar Persona

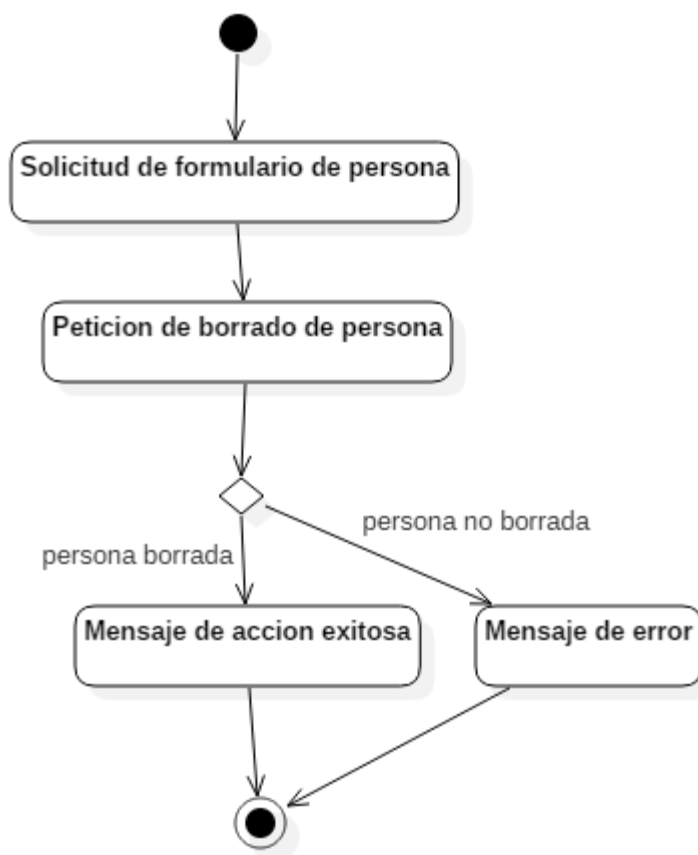


Figura 34. Diagrama de actividad Borrار Persona

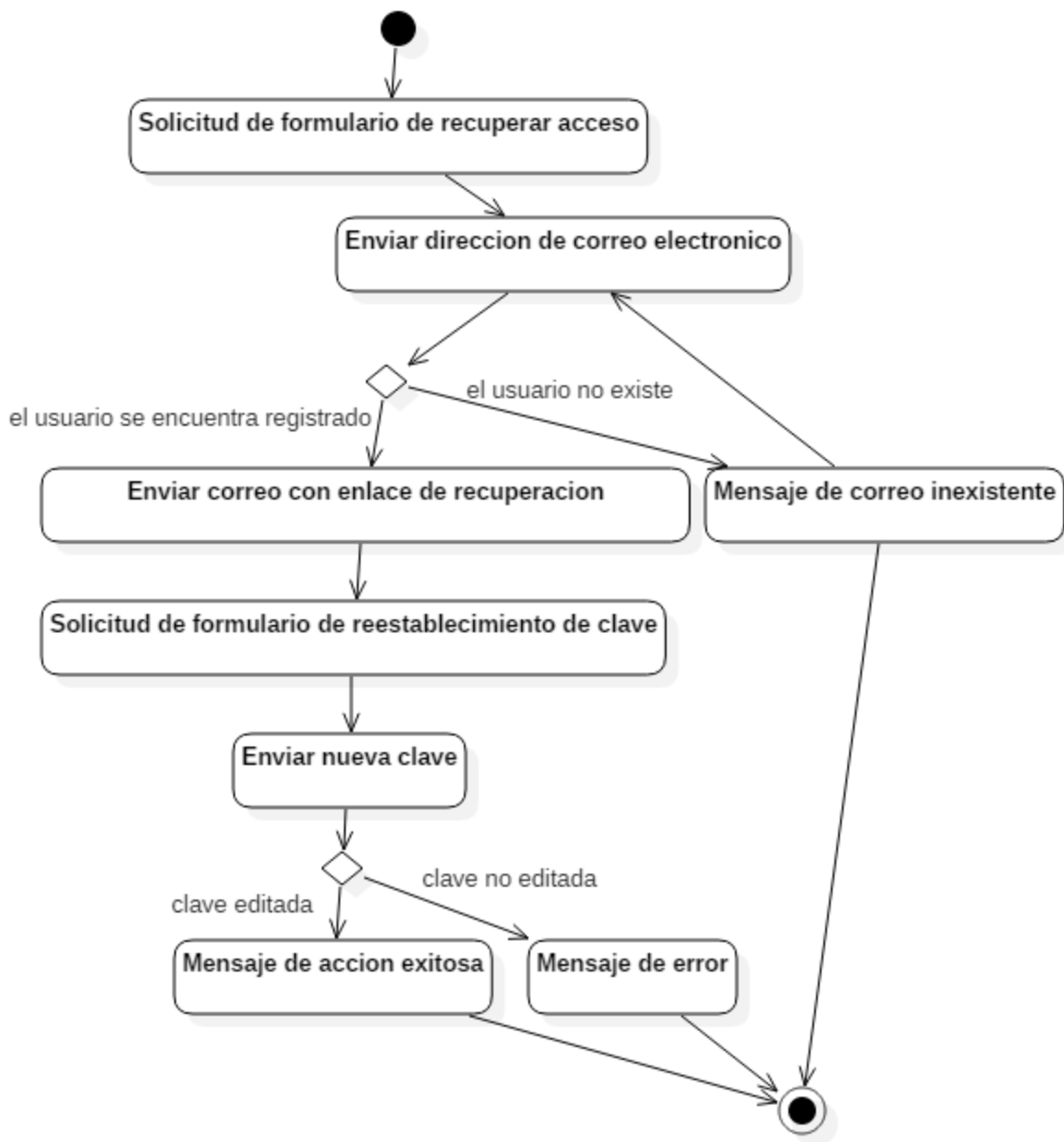


Figura 35. Diagrama de actividad Recuperar Acceso

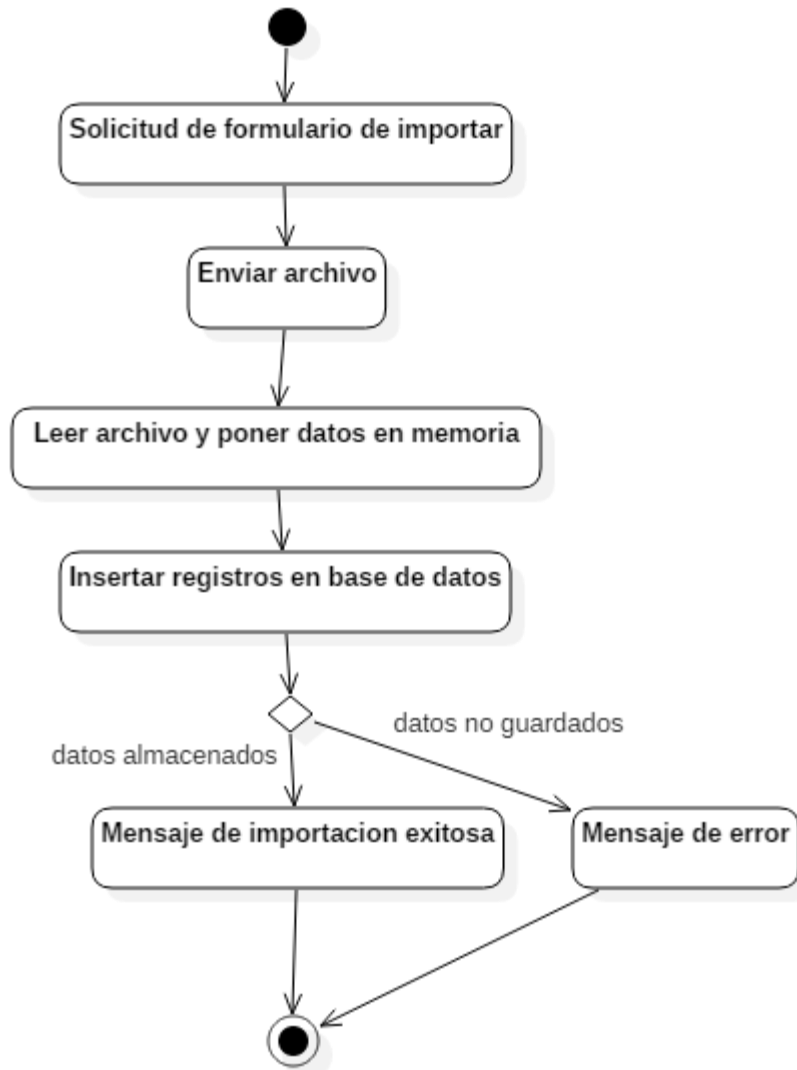


Figura 36. Diagrama de actividad Importar

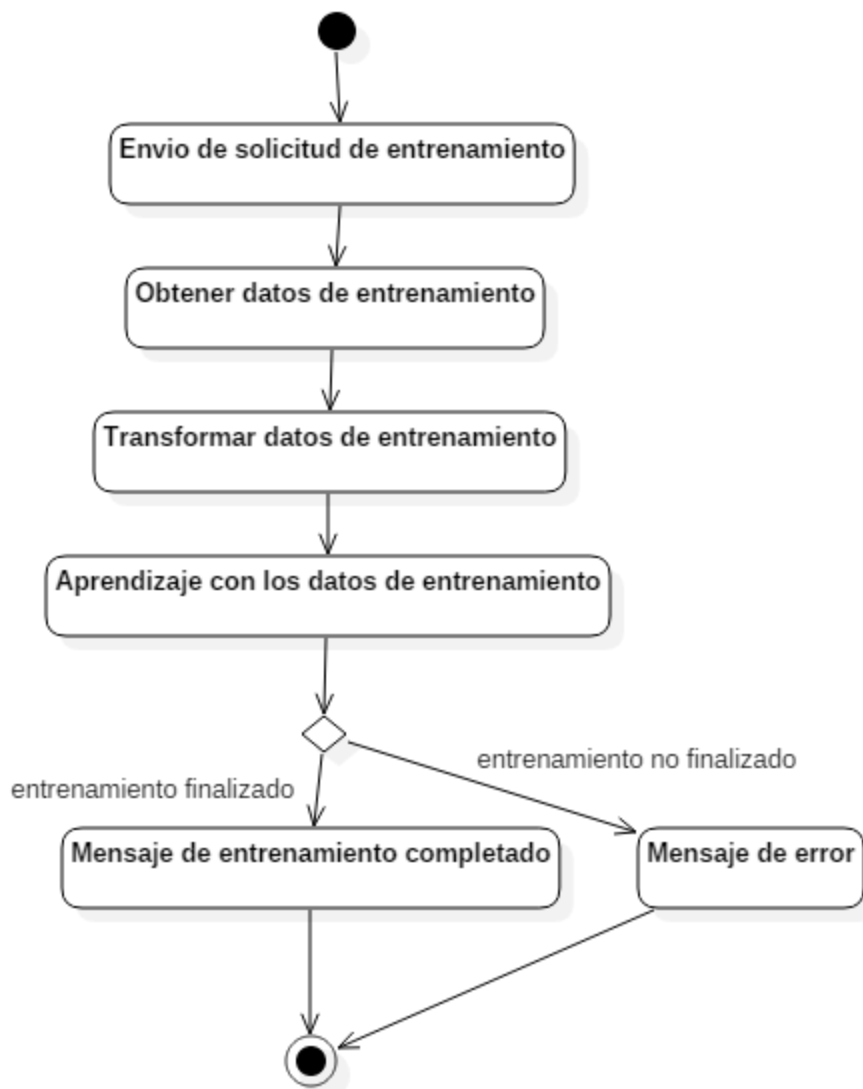


Figura 37. Diagrama de actividad Entrenar

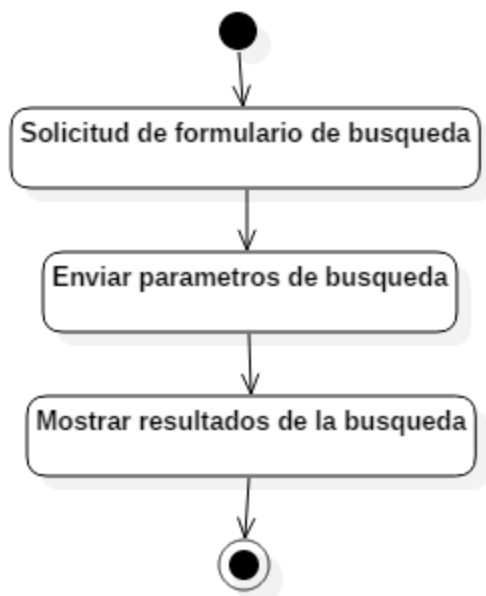


Figura 38. Diagrama de actividad Buscar

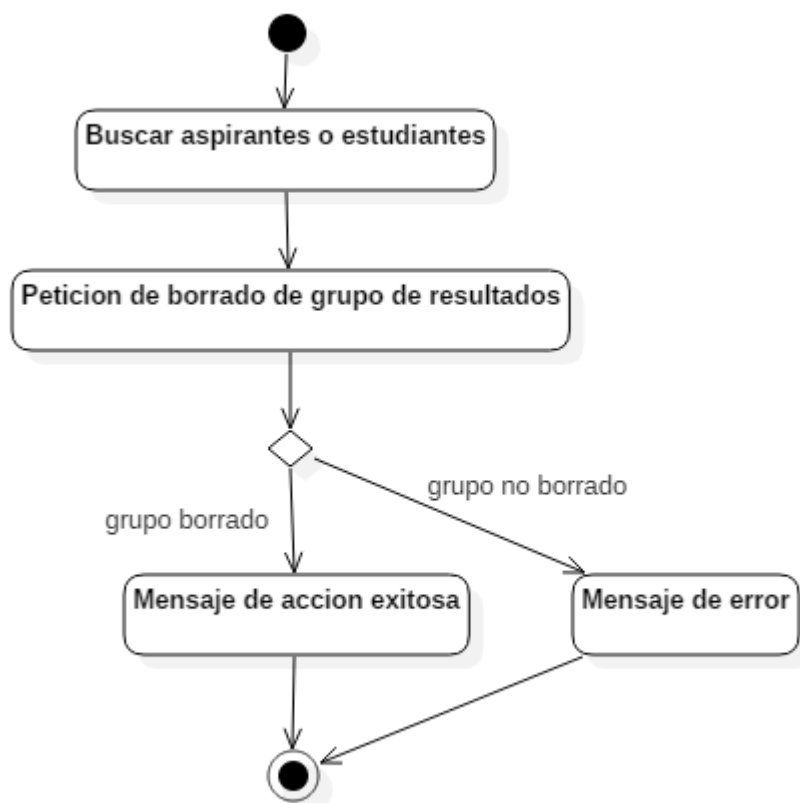


Figura 39. Diagrama de actividad Borrar Grupo

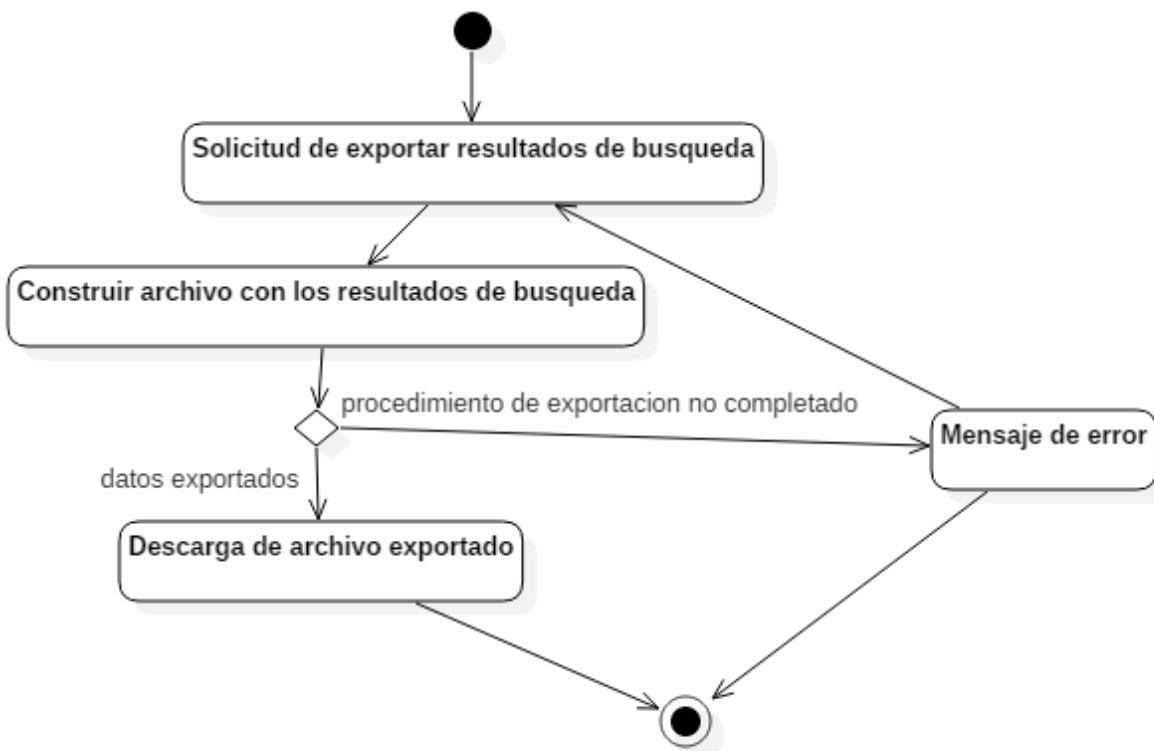


Figura 40. Diagrama de actividad Exportar

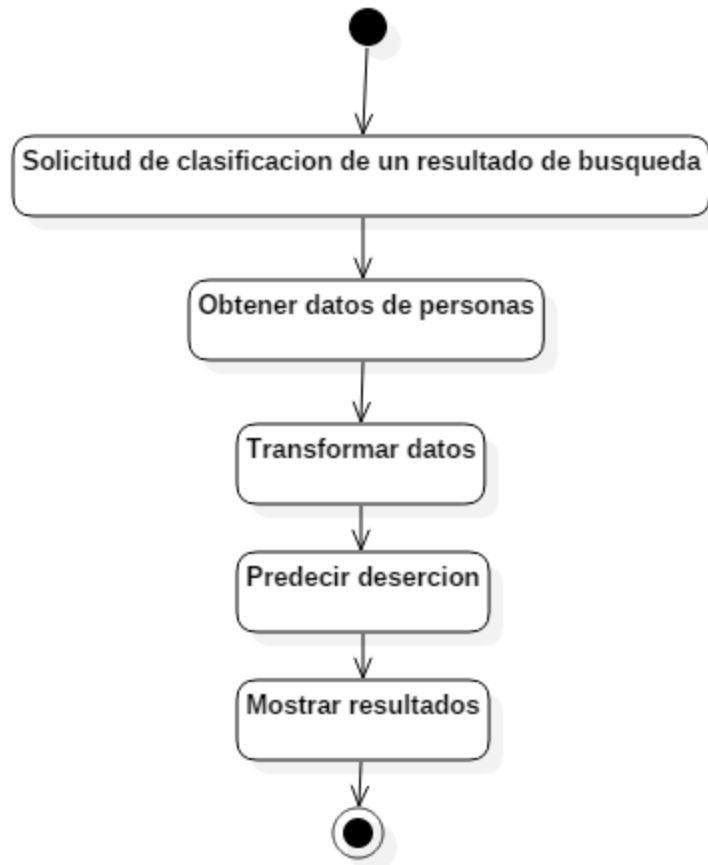


Figura 41. Diagrama de actividad Clasificar Grupo

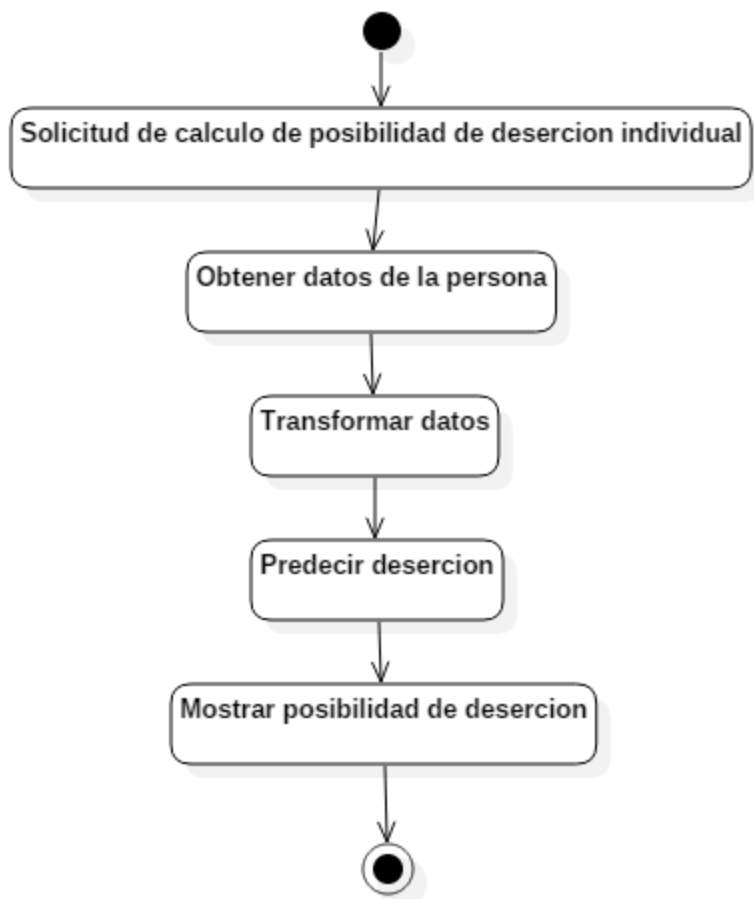


Figura 42. Diagrama de actividad Deserción Individual

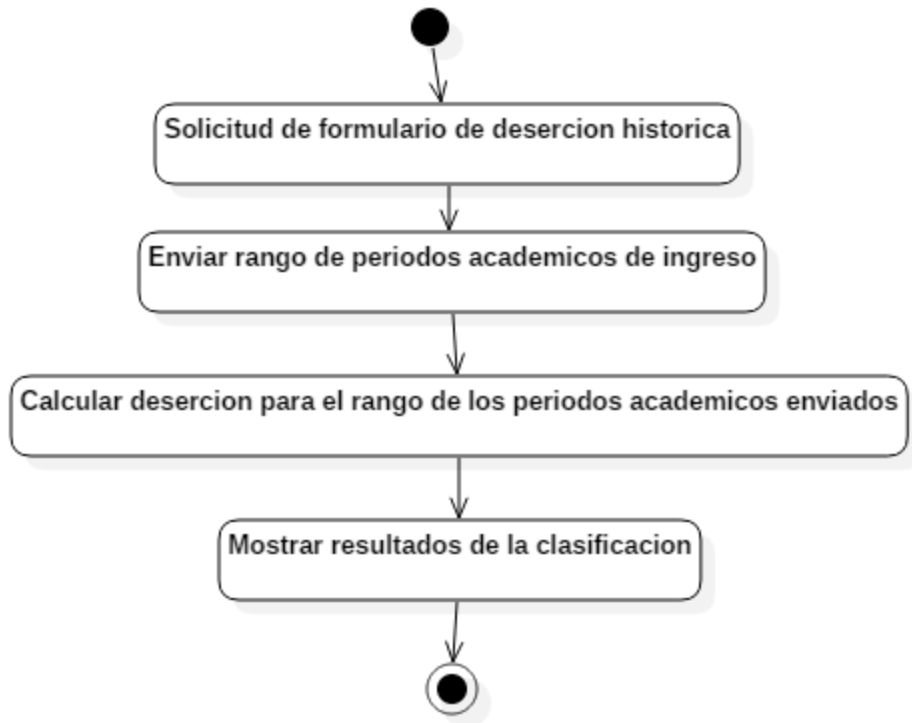


Figura 43. Diagrama de actividad Deserción Histórica

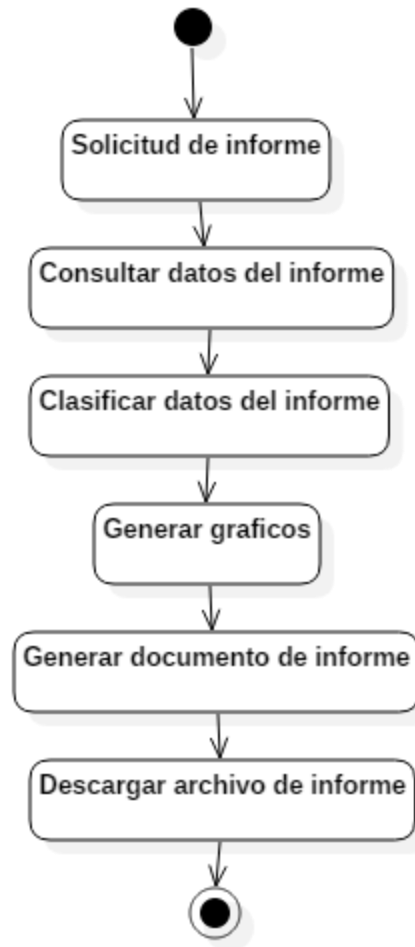


Figura 44. Diagrama de actividad Informe

Diagrama de Clases

El diagrama de clases describe la estructura del sistema mostrando las clases del mismo, así como sus funciones y atributos, además muestra la interacción entre clases. En la Figura 45 está modelado el diagrama de clases de la plataforma web.

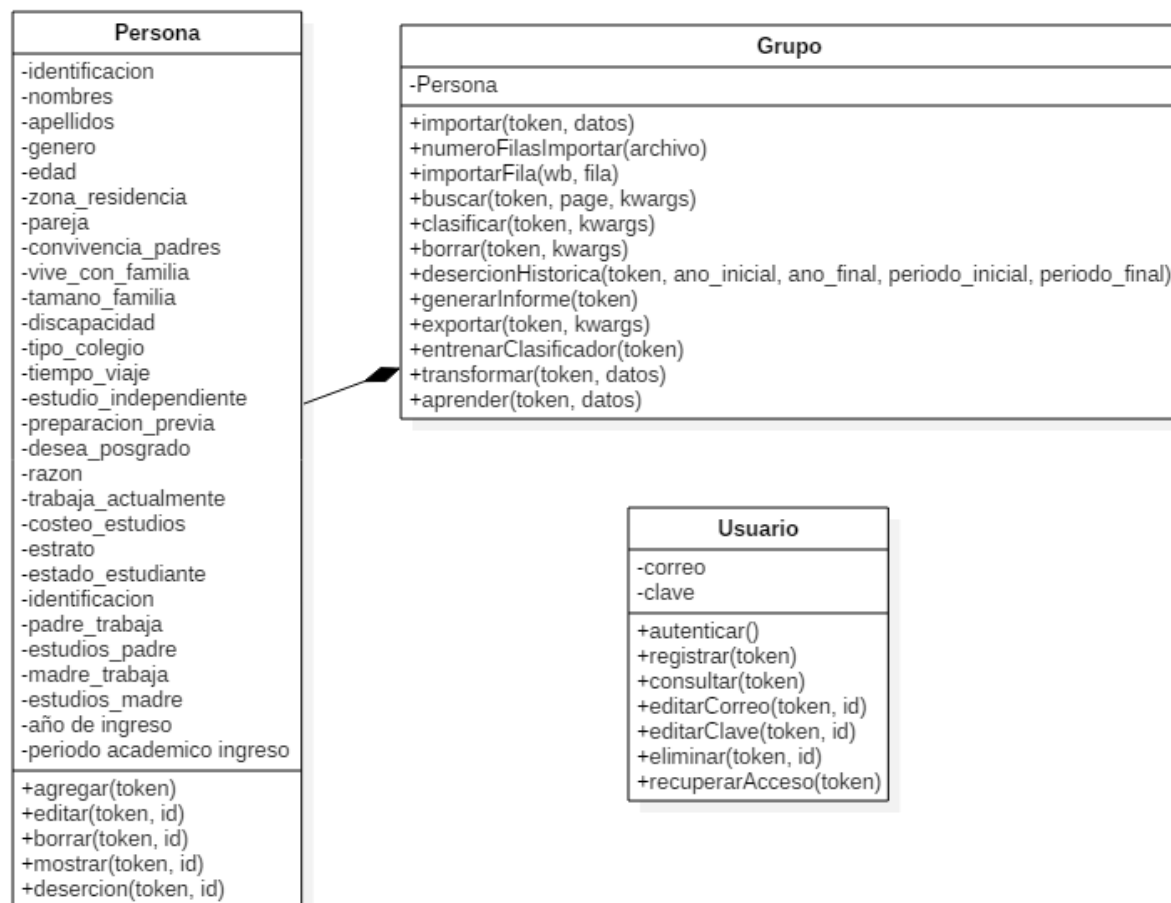


Figura 45. Diagrama de clases

Descripción del diagrama de clases

En la Tabla 12 se describen los atributos, funciones e interacciones de cada una de las clases del diagrama.

Tabla 12
Descripción del diagrama de clases

Clase	Descripción
Persona	Esta clase contiene todos los atributos de un aspirante o estudiante que son relevantes para el proceso de minería de datos y que sirven para identificar y agrupar los perfiles característicos, los métodos corresponden a los casos de uso que aplican para esta clase.

Grupo	Esta clase contiene un objeto como atributo, éste objeto es la composición de una o varias personas dentro de la case Grupo. Los métodos contienen las instrucciones de los casos de uso que aplican para esta clase.
Usuario	Esta clase encapsula los datos de acceso de un usuario, las funciones corresponden a las operaciones permitidas para un usuario de la plataforma web.

Dimensión organizacional

Diagrama de despliegue

El Diagrama de despliegue muestra la configuración de instalación recomendada de componentes de hardware y software para que la plataforma funcione correctamente.

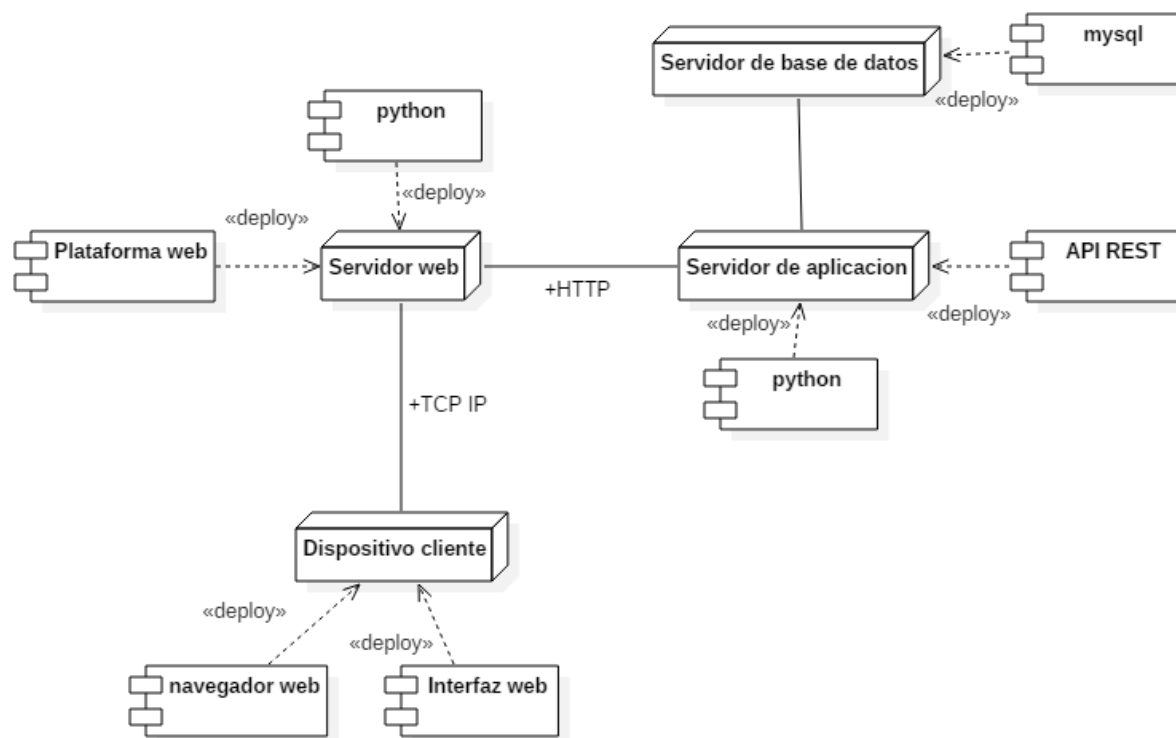


Figura 46. Diagrama de despliegue

Descripción del diagrama de despliegue

En la Tabla 13 se describen todos los elementos incluidos en el diagrama de despliegue.

Tabla 13
Descripción del diagrama de despliegue

Elemento	Descripción
Dispositivo cliente	Es el dispositivo de hardware desde donde accede el usuario de la plataforma web, se requiere un navegador web para acceder a la interfaz web.
Servidor web	Es un elemento de hardware que debe tener instalado el intérprete del lenguaje de programación python y los archivos de código fuente de la plataforma web.
Servidor de aplicación	Es un dispositivo electrónico donde se instala el código fuente del API REST y un intérprete de python para ejecutarlo, el servidor de aplicación puede ser el mismo servidor web pero requiere de la implementación del API REST en un puerto virtual diferente al de la plataforma web.
Servidor de base de datos	En el servidor de base de datos se instala el sistema gestor de base de datos mysql y se instala la base de datos de los servicios REST, el API REST crea la base de datos y si se desea se pueden importar datos de un archivo SQL para poblar las tablas.

Diseño de los casos de prueba

A continuación, se especifican los casos de prueba que se plantean para la plataforma web

MIDUDEC.

Tabla 14
Caso de prueba iniciar sesión en la plataforma web

Descripción	Iniciar sesión en la plataforma web.
Procedimiento	<ul style="list-style-type: none"> • Ingresar a la página web. • Diligenciar los datos de sesión. • Pulsar el botón Acceder.
Datos de prueba	Dirección de correo electrónico.
	Clave de acceso.

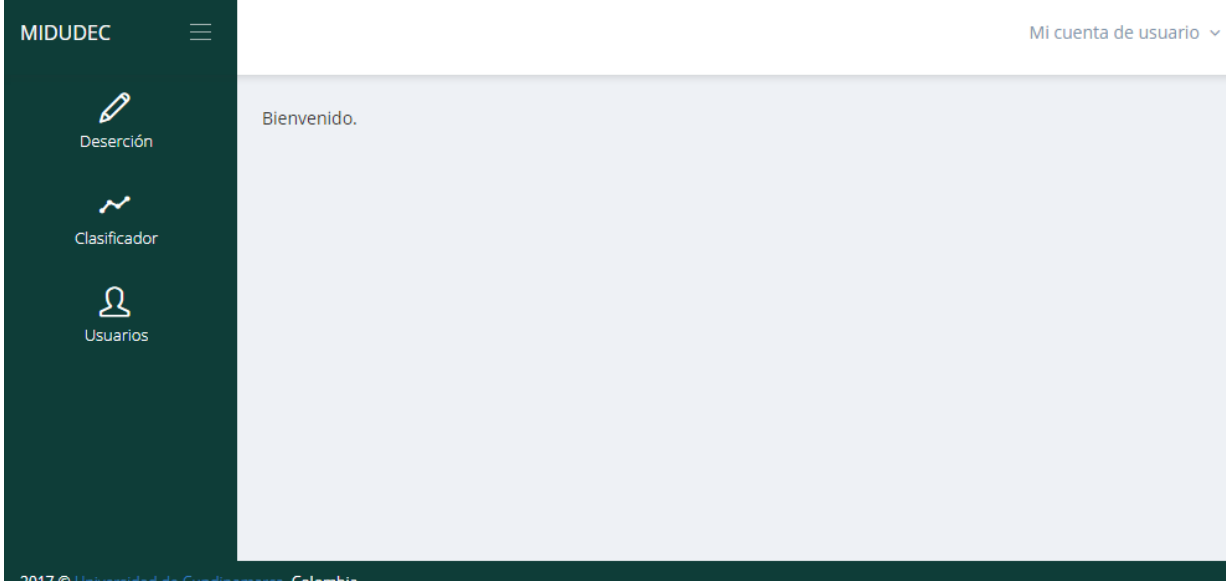
Resultados esperados	Si el usuario existe y la clave ingresada es correcta, el usuario tendrá acceso a su sesión.
Resultados obtenidos	Un usuario existente ingresó correctamente a la aplicación y fue redirigido a la interfaz principal de docente administrador.
 <p>The screenshot shows the MIDUDEC application interface. On the left is a dark green sidebar with the MIDUDEC logo and three menu items: 'Deserción' (with a pencil icon), 'Clasificador' (with a line graph icon), and 'Usuarios' (with a person icon). The main content area is light blue and displays 'Bienvenido.' at the top. In the top right corner, there is a link for 'Mi cuenta de usuario' with a dropdown arrow. At the bottom left of the interface, there is a copyright notice: '2017 © Universidad de Cundinamarca, Colombia'.</p>	
<p>Figura 47. Resultado de caso de prueba inicio de sesión</p>	

Tabla 15

Caso de prueba registrar un usuario docente administrador en la aplicación

Descripción	Registrar un usuario docente administrador en la aplicación.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Usuarios. • Seleccionar la opción Registrar usuarios. • Diligenciar los datos del formulario de registro. • Pulsar el botón registrar.
Datos de prueba	<p>Dirección de correo electrónico.</p> <p>Clave de acceso para el nuevo usuario.</p>
Resultados esperados	Si todas las casillas del formulario se han diligenciado correctamente se mostrará un aviso notificando la correcta creación del usuario.
Resultados obtenidos	La aplicación muestra un mensaje notificando que el nuevo usuario ha sido registrado exitosamente.



Figura 48. Resultado de caso de prueba registro de usuario

Tabla 16

Caso de prueba modificar los datos de acceso de un usuario autenticado en la aplicación

Descripción	Modificar los datos de acceso de un usuario autenticado en la aplicación.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar el menú desplegable Mi cuenta de usuario. • Seleccionar la opción Configurar perfil. • Escribir la nueva dirección de correo electrónico o la nueva clave que se desea establecer. • Pulsar el botón Cambiar correo o Cambiar clave según el dato que se desea modificar.
Datos de prueba	Dirección de correo electrónico. Clave de acceso.
Resultados esperados	Se mostrará un mensaje indicando el dato que se ha modificado.
Resultados obtenidos	La aplicación muestra un mensaje notificando que la dirección de correo electrónico se ha modificado. La aplicación muestra un mensaje notificando que la clave de acceso se ha modificado.

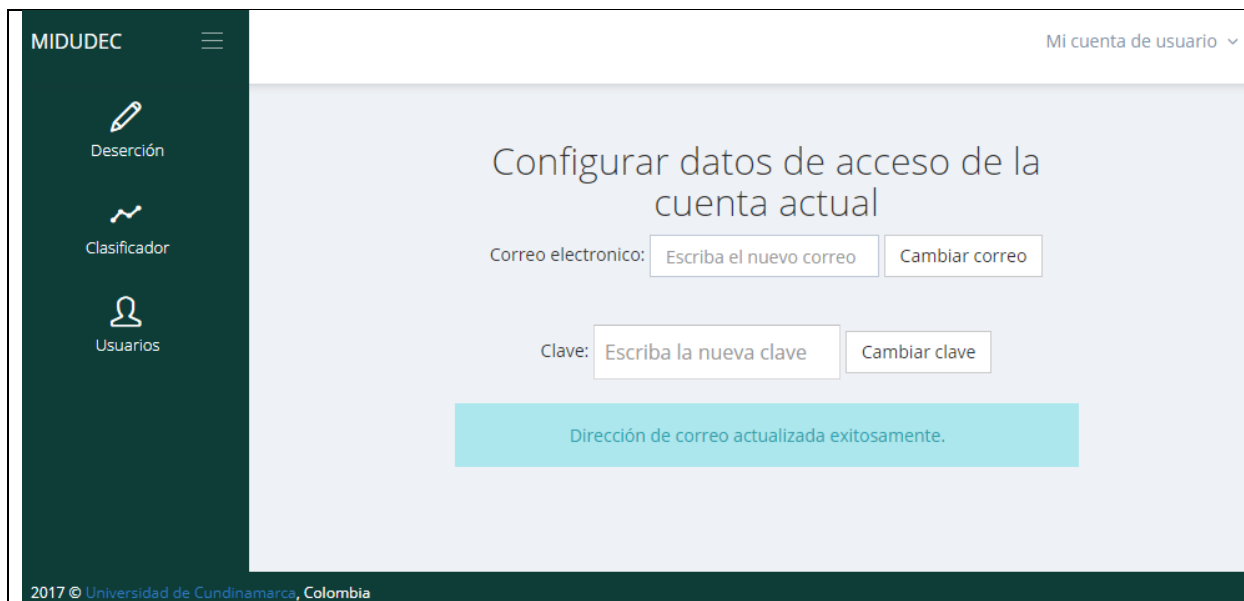


Figura 49. Resultado de caso de prueba configurar correo



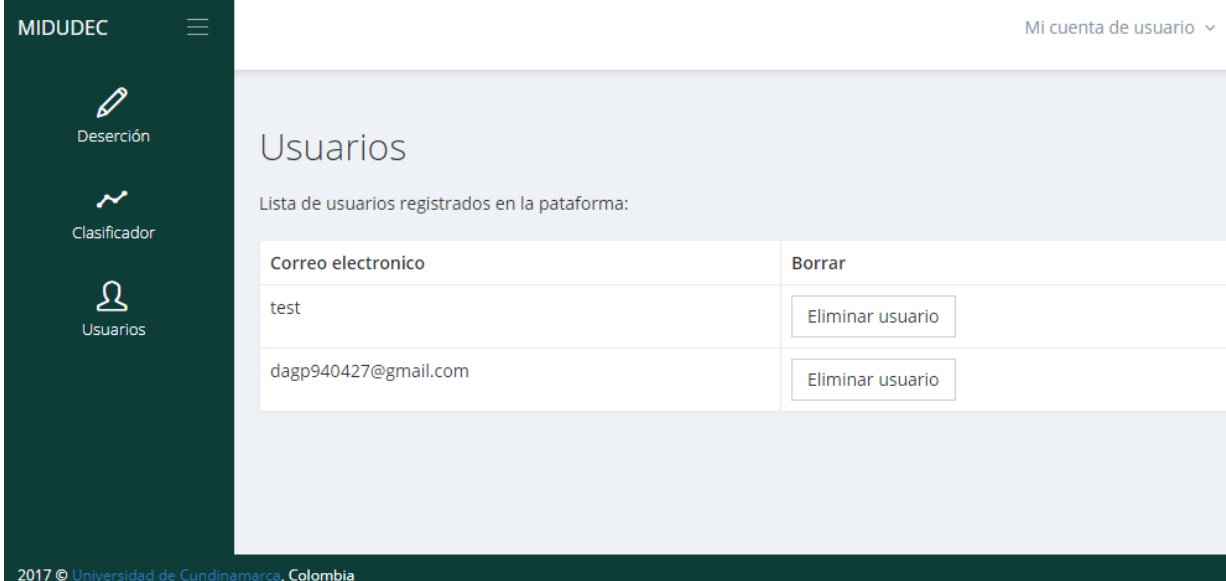
Figura 50. Resultado de caso de prueba configurar correo

Tabla 17

Caso de prueba consultar usuarios registrados en la plataforma web

Descripción	Consultar usuarios registrados en la plataforma web.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Usuarios. • Seleccionar la opción Consultar usuarios.

Resultados esperados	Una página con un listado de los usuarios registrados en el momento de la solicitud.
Resultados obtenidos	Se evidencia una tabla con todos los usuarios registrados en el momento de la realización de la prueba.



2017 © Universidad de Cundinamarca, Colombia

Figura 51. Resultado de caso de prueba consulta de usuarios

Tabla 18

Caso de prueba eliminar una cuenta de usuario de la plataforma web

Descripción	Eliminar una cuenta de usuario de la plataforma web.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Usuarios. • Seleccionar la opción Consultar usuarios. • Oprimir el botón Eliminar usuario que se encuentra en la misma fila de la cuenta de usuario que se requiere eliminar.
Resultados esperados	La cuenta de usuario es eliminada y desaparece de la lista de cuentas de usuario.
Resultados obtenidos	Se ha eliminado una cuenta de usuario, el usuario desaparece de la lista de consulta de usuarios y aparece un mensaje notificando que la operación se realizó con éxito.



Figura 52. Resultado de caso de prueba eliminación de usuarios

Tabla 19

Caso de prueba recuperar el acceso a la aplicación usando un mensaje de correo electrónico

Descripción	Recuperar el acceso a la aplicación usando un mensaje de correo electrónico.
Procedimiento	<ul style="list-style-type: none"> • Ingresar a la plataforma web. • Pulsar en el enlace Recuperar clave. • Escribir el correo electrónico de la cuenta de usuario. • Oprimir el botón Recuperar clave. • Ingresar al correo electrónico e ir al enlace que aparece en el mensaje enviado. • Escribir la nueva clave y oprimir el botón Cambiar clave.
Datos de prueba	Dirección de correo electrónico. Nueva clave de acceso.
Resultados esperados	Se enviará un mensaje con un enlace de cambio de clave con expiración al correo electrónico del usuario, luego se podrá establecer una nueva clave desde el enlace.
Resultados obtenidos	Se encontró un mensaje de correo electrónico en la bandeja de entrada, al acceder al enlace del mensaje se ingresa a un formulario donde se pudo

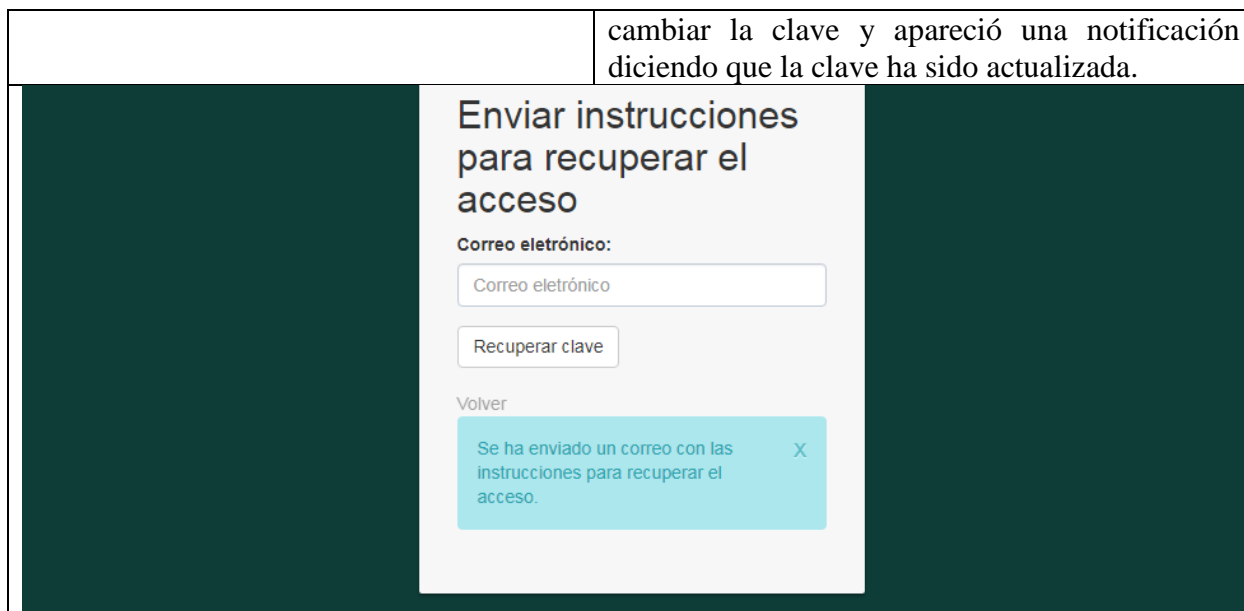


Figura 53. Resultado de caso de prueba enviar instrucciones de recuperación

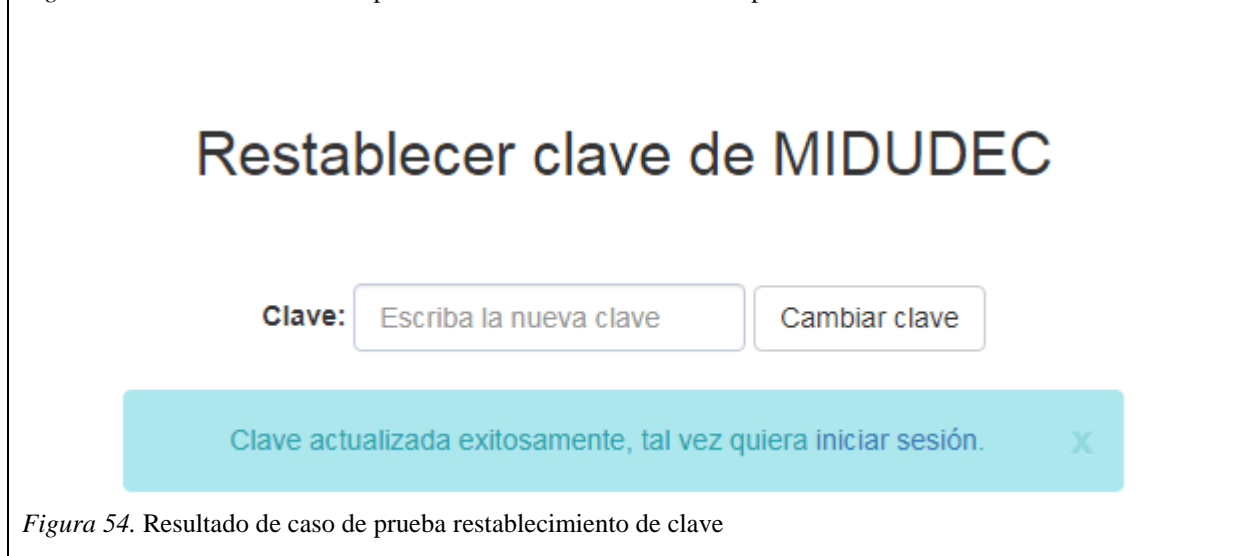


Figura 54. Resultado de caso de prueba restablecimiento de clave

Tabla 20

Caso de prueba agregar un aspirante o estudiante

Descripción	Agregar un aspirante o estudiante.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Agregar persona. • Diligenciar los datos del formulario. • Pulsar el botón agregar.
Datos de prueba	Número de identificación de la persona.

	Tipo de identificación.
	Nombres.
	Apellidos.
	Año de ingreso.
	Periodo académico de ingreso.
	Género.
	Fecha de nacimiento.
	Zona de residencia.
	Tiene pareja actualmente.
	Los padres conviven.
	El papá trabaja.
	La mamá trabaja.
	Nivel de estudios del padre.
	Nivel de estudios de la madre.
	Vive con la familia.
	Está en condición de discapacidad.
	Tipo de colegio donde estudió.
	Tiempo de viaje.
	Tiempo de estudio independiente.
	Preparación previa.
	Deseo de posgrado.
	Razón por la cual eligió la universidad.
	Trabaja actualmente.
	Financiamiento de los estudios.
	Estrato.
	Estado del estudiante.
Resultados esperados	Si todas las casillas del formulario se han diligenciado correctamente se mostrará un aviso notificando la correcta creación del aspirante o estudiante.
Resultados obtenidos	La aplicación muestra un mensaje notificando que la nueva persona ha sido agregada exitosamente.

MIDUDEC Mi cuenta de usuario ▾

Deserción
Clasificador
Usuarios

Persona creada exitosamente. Quizás quiera ver la persona agregada.

Registrar aspirante o estudiante

Número de identificación de la persona:
1098867343

Tipo de identificación:
C.C.

Nombres:
Pedro

Apellidos:
PÁRRIZ

Figura 55. Resultado de caso de prueba registro de un aspirante o estudiante

Tabla 21

Caso de prueba actualizar los datos de un aspirante o estudiante

Descripción	Actualizar los datos de un aspirante o estudiante.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Seleccionar el enlace de la persona que se desea editar. • Diligenciar los campos del formulario que se desean modificar. • Pulsar el botón Guardar cambios.
Datos de prueba	Número de identificación de la persona.
	Tipo de identificación.
	Nombres.
	Apellidos.
	Año de ingreso.
	Periodo académico de ingreso.
	Género.
	Fecha de nacimiento.


	Zona de residencia.
	Tiene pareja actualmente.
	Los padres conviven.
	El papá trabaja.
	La mamá trabaja.
	Nivel de estudios del padre.
	Nivel de estudios de la madre.
	Vive con la familia.
	Está en condición de discapacidad.
	Tipo de colegio donde estudió.
	Tiempo de viaje.
	Tiempo de estudio independiente.
	Preparación previa.
	Deseo de posgrado.
	Razón por la cual eligió la universidad.
	Trabaja actualmente.
	Financiamiento de los estudios.
	Estrato.
	Estado del estudiante.
Resultados esperados	Si todas las casillas del formulario se han diligenciado correctamente se mostrará un aviso notificando la correcta edición del aspirante o estudiante.
Resultados obtenidos	Se muestra una alerta indicando que la persona se modificó correctamente.
 <p>The screenshot shows the MIDUDEC application interface. On the left is a dark green sidebar with navigation icons for 'Deserción', 'Clasificador', and 'Usuarios'. The main content area displays a light blue success message: 'Persona editada exitosamente.' Below this is the 'Datos de la persona' form, which includes a checkbox for '¿Es un posible desertor?', a text input for 'Número de identificación de la persona' containing '1098867343', a dropdown menu for 'Tipo de identificación' set to 'C.C.', and a text input for 'Nombres' containing 'Pedro'. The top right corner shows 'Mi cuenta de usuario' with a dropdown arrow.</p>	

Figura 56. Resultado de caso de prueba edición de un aspirante o estudiante

Tabla 22

Caso de prueba visualizar los datos de un aspirante o estudiante

Descripción	Visualizar los datos de un aspirante o estudiante.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Seleccionar el enlace de la persona que se desea visualizar.
Resultados esperados	Un formulario con los datos de un aspirante o estudiante.
Resultados obtenidos	Se accede a un formulario cuyos campos están rellenos con los datos de la persona.

MIDUDEC Mi cuenta de usuario ▾

Género: Masculino Femenino

Fecha de nacimiento:

Zona de residencia:

Urbana
 Rural

¿Tiene pareja?: No Si

¿Los padres conviven?: Si No

¿El papá trabaja?: Si No

¿La mamá trabaja?: Si No

Nivel de estudios del padre:

Figura 57. Resultado de caso de prueba visualización de un aspirante o estudiante

Tabla 23

Caso de prueba borrar un aspirante o estudiante

Descripción	Borrar un aspirante o estudiante.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes.

	<ul style="list-style-type: none"> • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Seleccionar el enlace de la persona que se desea borrar. • Pulsar el botón Borrar persona.
Resultados esperados	Se mostrará un aviso indicando si la persona se eliminó.
Resultados obtenidos	La plataforma se redirige al formulario de registro de personas y muestra un mensaje que dice: Datos de la persona eliminados exitosamente.



Figura 58. Resultado de caso de prueba eliminación de un aspirante o estudiante

Tabla 24

Caso de prueba calcular la posibilidad de deserción de una persona

Descripción	Calcular la posibilidad de deserción de una persona.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Seleccionar el enlace de una persona • Pulsar el botón ¿Es un posible desertor?

Resultados esperados	Se mostrará el texto si o no dependiendo de si la persona es un posible desertor o no lo es.
Resultados obtenidos	Se evidencia que la persona consultada no es un posible desertor, se muestra el texto “No” debajo del botón ¿Es un posible desertor?

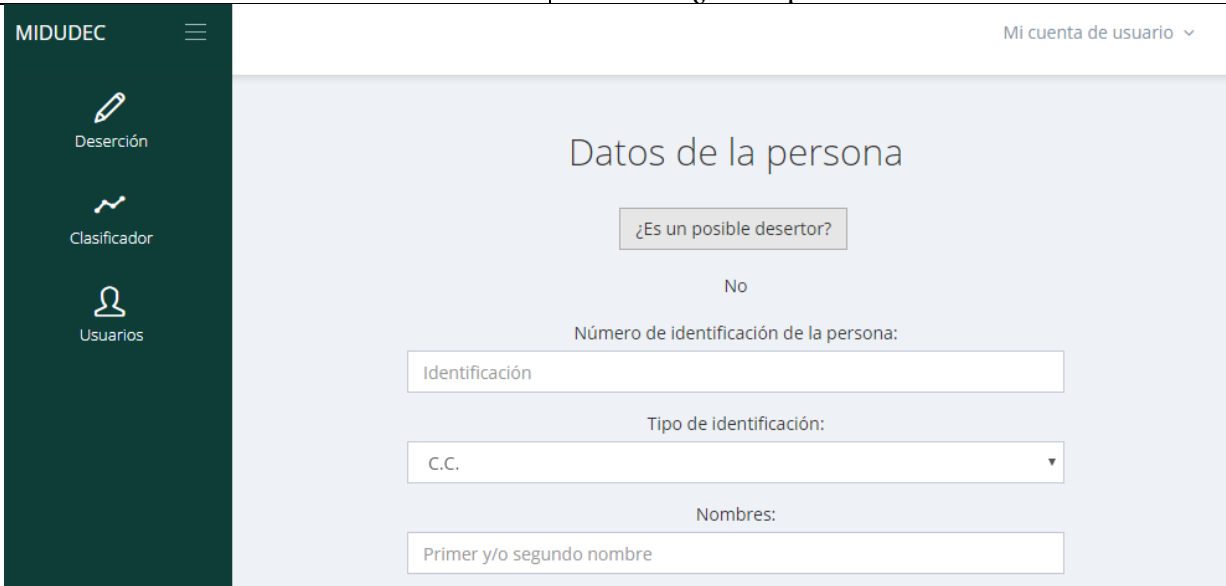


Figura 59. Resultado de caso de prueba determinación de la posibilidad de deserción de un aspirante o estudiante

Tabla 25
Caso de prueba importar datos de un grupo de personas

Descripción	Importar datos de un grupo de personas.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Importar grupo de personas. • Pulsar el botón de selección de archivo. • Seleccionar el archivo de hoja de cálculo con los datos de las personas a importar. • Pulsar el botón Cargar archivo.
Datos de prueba	Hoja de cálculo con datos de perfiles de estudiantes o aspirantes.
Resultados esperados	Perfiles de estudiantes importados desde la hoja de cálculo.
Resultados obtenidos	Se muestra una barra de progreso que indica el estado del proceso de importación, al completarse

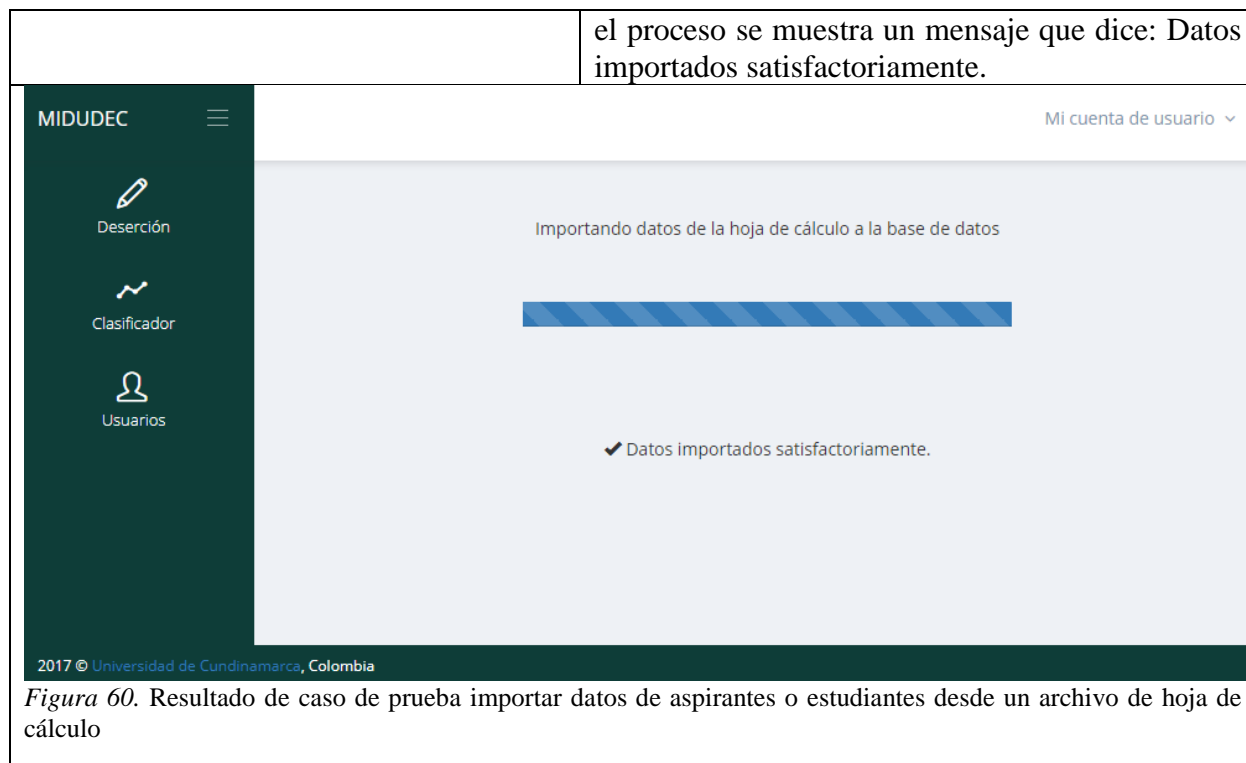


Tabla 26

Caso de prueba buscar aspirantes o estudiantes

Descripción	Buscar aspirantes o estudiantes.											
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. 											
Datos de prueba	<table border="1"> <tr><td>Número de identificación de la persona.</td></tr> <tr><td>Tipo de identificación.</td></tr> <tr><td>Nombres.</td></tr> <tr><td>Apellidos.</td></tr> <tr><td>Año de ingreso.</td></tr> <tr><td>Periodo académico de ingreso.</td></tr> <tr><td>Género.</td></tr> <tr><td>Fecha de nacimiento.</td></tr> <tr><td>Zona de residencia.</td></tr> <tr><td>Tiene pareja actualmente.</td></tr> <tr><td>Los padres conviven.</td></tr> </table>	Número de identificación de la persona.	Tipo de identificación.	Nombres.	Apellidos.	Año de ingreso.	Periodo académico de ingreso.	Género.	Fecha de nacimiento.	Zona de residencia.	Tiene pareja actualmente.	Los padres conviven.
Número de identificación de la persona.												
Tipo de identificación.												
Nombres.												
Apellidos.												
Año de ingreso.												
Periodo académico de ingreso.												
Género.												
Fecha de nacimiento.												
Zona de residencia.												
Tiene pareja actualmente.												
Los padres conviven.												

	<p>El papá trabaja.</p> <p>La mamá trabaja.</p> <p>Nivel de estudios del padre.</p> <p>Nivel de estudios de la madre.</p> <p>Vive con la familia.</p> <p>Está en condición de discapacidad.</p> <p>Tipo de colegio donde estudió.</p> <p>Tiempo de viaje.</p> <p>Tiempo de estudio independiente.</p> <p>Preparación previa.</p> <p>Deseo de posgrado.</p> <p>Razón por la cual eligió la universidad.</p> <p>Trabaja actualmente.</p> <p>Financiamiento de los estudios.</p> <p>Estrato.</p> <p>Estado del estudiante.</p>
Resultados esperados	Lista de personas cuyos datos coinciden con los filtros establecidos.
Resultados obtenidos	Se evidencia una página de resultados con enlaces a perfiles de personas cuyos datos corresponden a los filtros establecidos en la página de búsqueda.

Figura 61. Resultado de caso de prueba búsqueda de aspirantes o estudiantes

Tabla 27

Caso de prueba borrar un grupo de personas pertenecientes a un resultado de búsqueda

Descripción	Borrar un grupo de personas pertenecientes a un resultado de búsqueda.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Pulsar en el botón Borrar. • Seleccionar la opción “Si, borrar grupo” en el cuadro de confirmación.
Resultados esperados	Se mostrará un aviso indicando que se ha borrado el grupo de personas.
Resultados obtenidos	Se evidencia una ventana informando la eliminación de un grupo de personas, al presionar el botón OK, la plataforma se redirige a la página de búsqueda de personas.

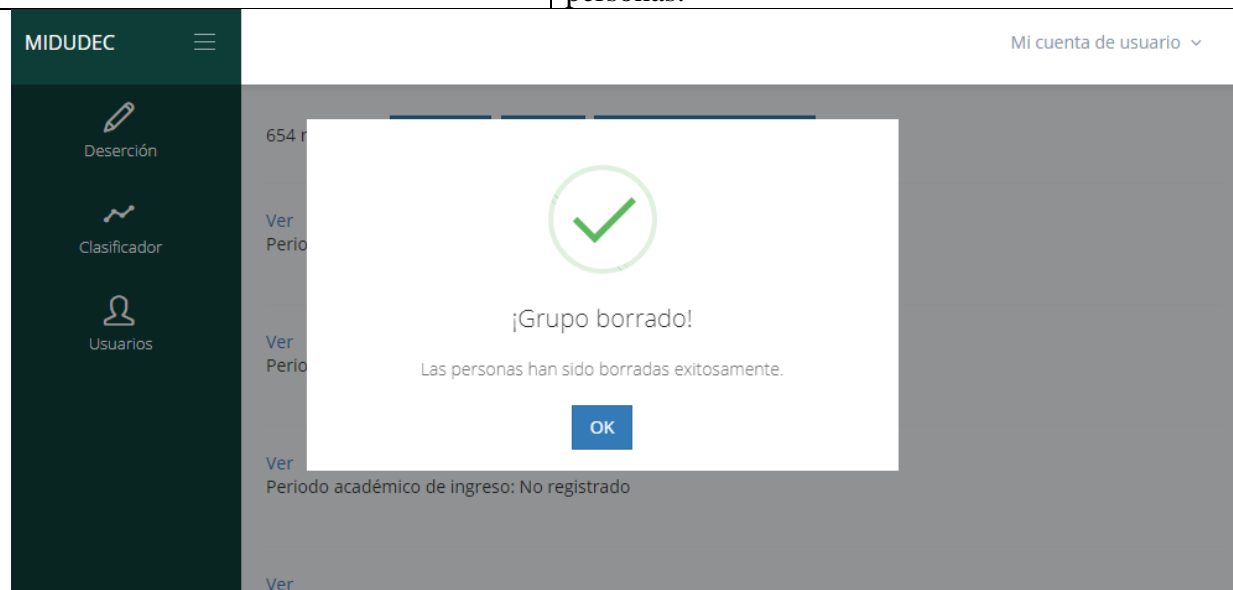


Figura 62. Resultado de caso de prueba eliminación de un grupo de aspirantes o estudiantes

Tabla 28

Caso de prueba entrenar el algoritmo de clasificación

Descripción	Entrenar el algoritmo de clasificación.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web.

	<ul style="list-style-type: none"> • Importar el grupo de personas que componen los datos de entrenamiento. • Seleccionar la categoría Clasificador. • Seleccionar la opción Entrenar algoritmo de minería de datos. • Presionar el botón Entrenar.
Datos de prueba	Hoja de cálculo con datos de perfiles de estudiantes o aspirantes.
Resultados esperados	Un mensaje notificando que se ha realizado el entrenamiento del modelo de minería de datos.
Resultados obtenidos	Se puede observar una barra de progreso que se llena a medida que se completa el entrenamiento, al finalizar aparece un texto que dice Entrenamiento finalizado y un enlace para recargar la página y ver los resultados del entrenamiento realizado, al seleccionar el enlace se recarga la página con las siguientes medidas de desempeño: exactitud, curva de aprendizaje, árbol de decisión y matriz de confusión.

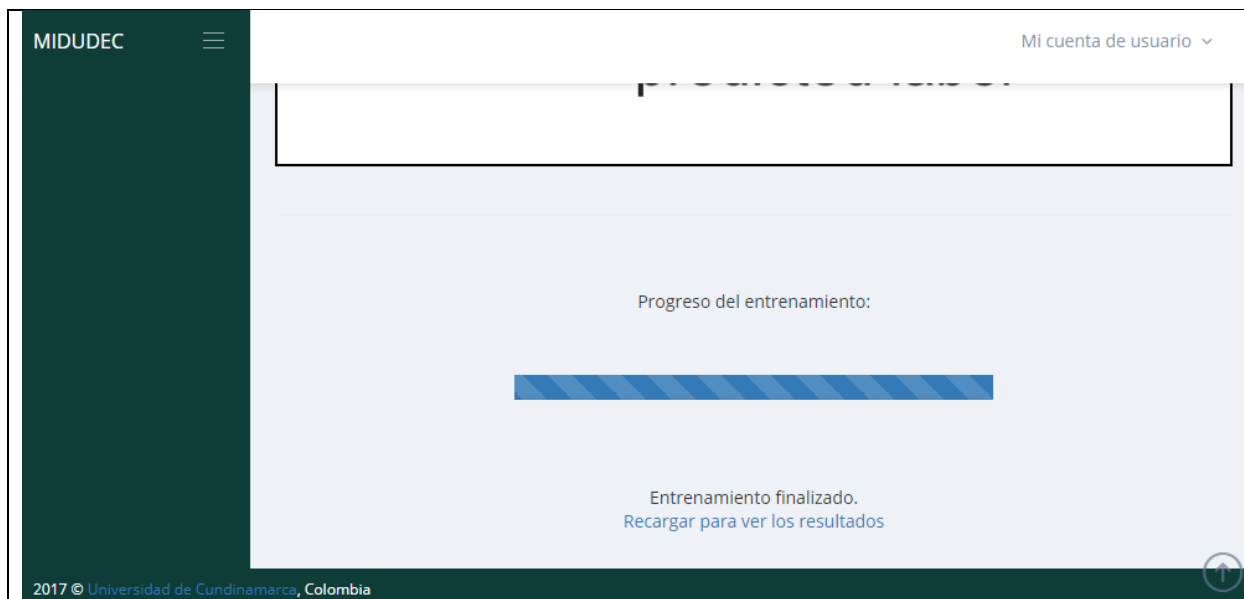


Figura 63. Resultado de caso de prueba entrenamiento del algoritmo de clasificación



Figura 64. Resultado de caso de prueba resultados de entrenamiento

Tabla 29

Caso de prueba clasificar un grupo de personas por posibilidad de deserción

Descripción	Clasificar un grupo de personas por
--------------------	-------------------------------------

	posibilidad de deserción.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Pulsar en el botón Clasificar por deserción.
Resultados esperados	Una página con una tabla y gráficas sobre la posibilidad de deserción de un grupo de personas.
Resultados obtenidos	Una tabla con la cantidad y porcentaje de posibles desertores y no desertores. Un gráfico de barras y un gráfico de torta con la cantidad de desertores y no desertores.

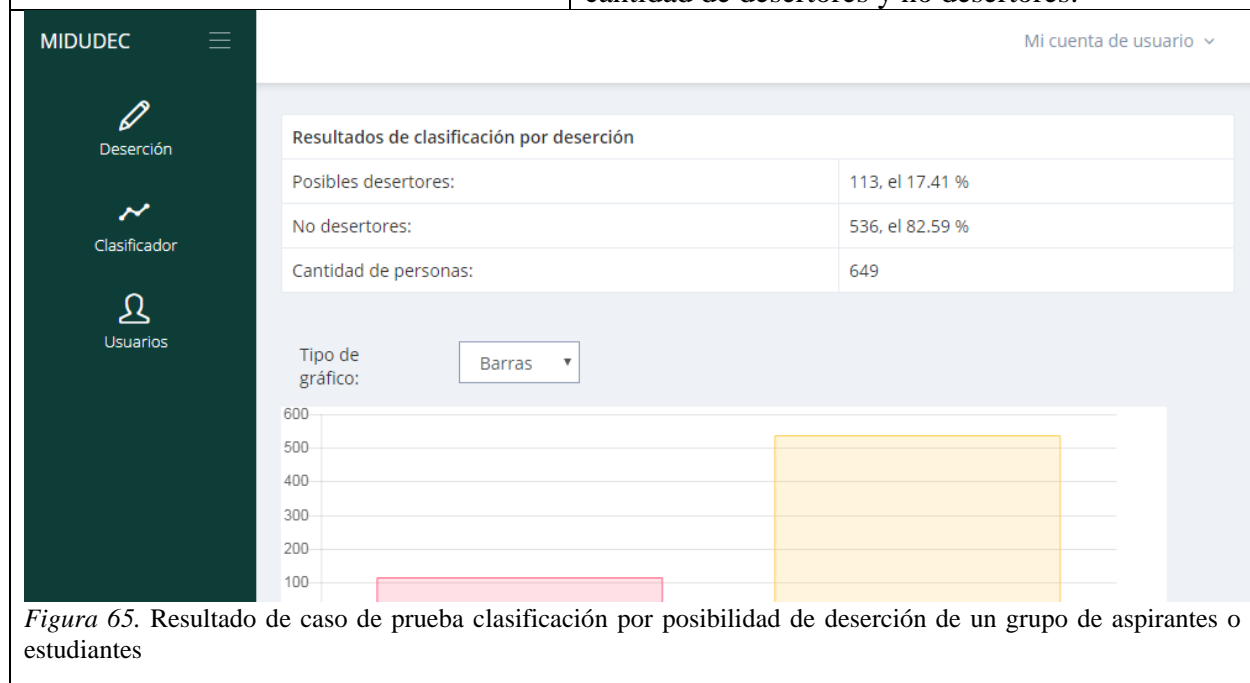


Figura 65. Resultado de caso de prueba clasificación por posibilidad de deserción de un grupo de aspirantes o estudiantes

Tabla 30

Caso de prueba calcular la posibilidad de deserción histórica

Descripción	Calcular la posibilidad de deserción histórica en un rango de periodos académicos de ingreso.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción.

	<ul style="list-style-type: none"> • Seleccionar la opción Deserción histórica. • Ingresar el rango de periodos académicos de ingreso deseados. • Pulsar el botón Calcular deserción.
Datos de prueba	Año de ingreso inicial.
	Periodo académico de ingreso inicial.
	Año de ingreso final.
	Periodo académico de ingreso final.
Resultados esperados	Una tabla con la clasificación de personas por deserción que pertenecen al rango de periodos académicos de ingreso establecidos y una gráfica describiendo el fenómeno en cada cohorte.
Resultados obtenidos	Una tabla con la cantidad y porcentaje de posibles desertores y no desertores para los periodos académicos de ingreso establecidos. Una gráfica de línea de tiempo con la cantidad de desertores y no desertores en cada periodo académico de ingreso dentro del rango establecido.

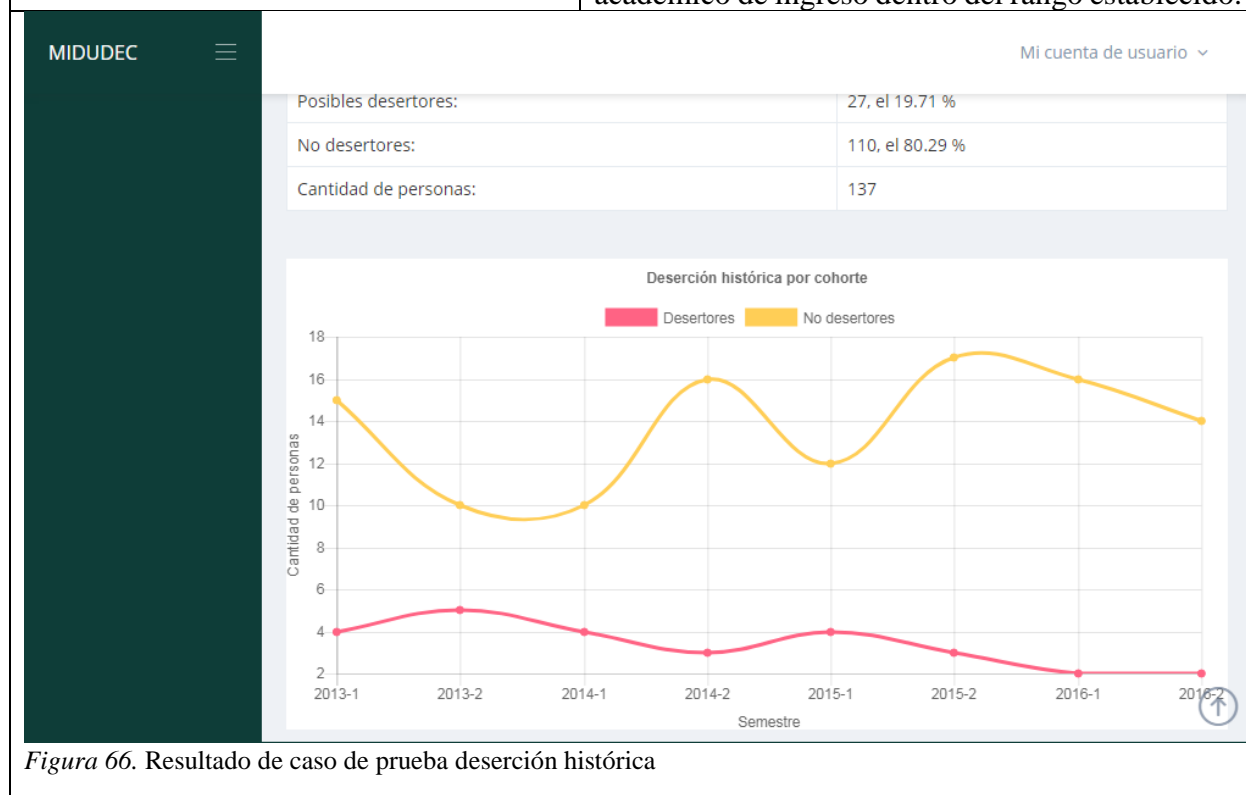


Figura 66. Resultado de caso de prueba deserción histórica

Tabla 31

Caso de prueba exportar datos de perfiles de personas a un archivo de hoja de cálculo

Descripción	Exportar datos de perfiles de personas a un archivo de hoja de cálculo.																																																																																																																																																																																																																																																																																																																																				
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Buscar aspirantes o estudiantes. • Establecer los filtros necesarios para encontrar un grupo de personas deseado. • Presionar el botón Buscar. • Pulsar el botón Exportar. 																																																																																																																																																																																																																																																																																																																																				
Resultados esperados	Un archivo de hoja de cálculo con los datos de un grupo de personas.																																																																																																																																																																																																																																																																																																																																				
Resultados obtenidos	Se genera un archivo de hoja de cálculo para descargar.																																																																																																																																																																																																																																																																																																																																				
<table border="1"> <thead> <tr> <th>ano_ingr</th> <th>periodo</th> <th>genero</th> <th>fecha_nacimiento</th> <th>zona_res</th> <th>pareja</th> <th>padres_c</th> <th>padre_tri</th> <th>madre_tri</th> <th>estudios</th> <th>estudios</th> <th>vive_con</th> <th>tamano</th> <th>discapac</th> <th>tipo_col</th> <th>tiempo</th> <th>v_estudio</th> <th>prepar</th> </tr> </thead> <tbody> <tr> <td>2011</td> <td>1</td> <td>M</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Media</td> <td>Si</td> <td>Menor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2005</td> <td>2</td> <td>F</td> <td>2000-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Separad</td> <td>Si</td> <td>Si</td> <td>Profesior</td> <td>Profesior</td> <td>Si</td> <td>Mayor a</td> <td>Si</td> <td>Privado</td> <td>15 a 30 m</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2016</td> <td>1</td> <td>M</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Separad</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Secundai</td> <td>Si</td> <td>Menor o</td> <td>Si</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2012</td> <td>2</td> <td>F</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Profesior</td> <td>Profesior</td> <td>Si</td> <td>Mayor a</td> <td>Si</td> <td>Privado</td> <td>Menor a</td> <td>Menor a</td> <td>No</td> </tr> <tr> <td>2013</td> <td>2</td> <td>F</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Media</td> <td>No</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>Menor a</td> <td>No</td> </tr> <tr> <td>2010</td> <td>1</td> <td>M</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Media</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>Menor a</td> <td>No</td> </tr> <tr> <td>2003</td> <td>2</td> <td>M</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Profesior</td> <td>Profesior</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>Si</td> </tr> <tr> <td>2006</td> <td>1</td> <td>M</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Secundai</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>Menor a</td> <td>No</td> </tr> <tr> <td>2008</td> <td>1</td> <td>F</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Separad</td> <td>Si</td> <td>Si</td> <td>Secundai</td> <td>Secundai</td> <td>No</td> <td>Menor o</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2009</td> <td>2</td> <td>M</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Profesior</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2008</td> <td>2</td> <td>M</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Secundai</td> <td>Media</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>5 a 10 ho</td> <td>No</td> </tr> <tr> <td>2013</td> <td>1</td> <td>F</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>Si</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Secundai</td> <td>Secundai</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>5 a 10 ho</td> <td>No</td> </tr> <tr> <td>2013</td> <td>2</td> <td>F</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Secundai</td> <td>Si</td> <td>Mayor a</td> <td>Si</td> <td>Privado</td> <td>15 a 30 m</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2015</td> <td>1</td> <td>F</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>Si</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Media</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2009</td> <td>2</td> <td>F</td> <td>2001-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Separad</td> <td>Si</td> <td>Si</td> <td>Media</td> <td>Media</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>Menor a</td> <td>No</td> </tr> <tr> <td>2000</td> <td>2</td> <td>F</td> <td>2002-04-19 0:00:00</td> <td>U</td> <td>Si</td> <td>Separad</td> <td>Si</td> <td>Si</td> <td>Secundai</td> <td>Secundai</td> <td>No</td> <td>Mayor a</td> <td>Si</td> <td>Privado</td> <td>Menor a</td> <td>2 a 5 hori</td> <td>No</td> </tr> <tr> <td>2001</td> <td>2</td> <td>M</td> <td>2000-04-19 0:00:00</td> <td>U</td> <td>No</td> <td>Juntos</td> <td>Si</td> <td>Si</td> <td>Primaria</td> <td>Media</td> <td>Si</td> <td>Mayor a</td> <td>No</td> <td>Privado</td> <td>Menor a</td> <td>Menor a</td> <td>No</td> </tr> </tbody> </table>		ano_ingr	periodo	genero	fecha_nacimiento	zona_res	pareja	padres_c	padre_tri	madre_tri	estudios	estudios	vive_con	tamano	discapac	tipo_col	tiempo	v_estudio	prepar	2011	1	M	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Media	Si	Menor a	No	Privado	Menor a	2 a 5 hori	No	2005	2	F	2000-04-19 0:00:00	U	No	Separad	Si	Si	Profesior	Profesior	Si	Mayor a	Si	Privado	15 a 30 m	2 a 5 hori	No	2016	1	M	2002-04-19 0:00:00	U	No	Separad	Si	Si	Media	Secundai	Si	Menor o	Si	Privado	Menor a	2 a 5 hori	No	2012	2	F	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Profesior	Profesior	Si	Mayor a	Si	Privado	Menor a	Menor a	No	2013	2	F	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Media	No	Mayor a	No	Privado	Menor a	Menor a	No	2010	1	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Media	Si	Mayor a	No	Privado	Menor a	Menor a	No	2003	2	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Profesior	Profesior	Si	Mayor a	No	Privado	Menor a	2 a 5 hori	Si	2006	1	M	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Secundai	Si	Mayor a	No	Privado	Menor a	Menor a	No	2008	1	F	2001-04-19 0:00:00	U	No	Separad	Si	Si	Secundai	Secundai	No	Menor o	No	Privado	Menor a	2 a 5 hori	No	2009	2	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Profesior	Si	Mayor a	No	Privado	Menor a	2 a 5 hori	No	2008	2	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Secundai	Media	Si	Mayor a	No	Privado	Menor a	5 a 10 ho	No	2013	1	F	2001-04-19 0:00:00	U	Si	Juntos	Si	Si	Secundai	Secundai	Si	Mayor a	No	Privado	Menor a	5 a 10 ho	No	2013	2	F	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Secundai	Si	Mayor a	Si	Privado	15 a 30 m	2 a 5 hori	No	2015	1	F	2001-04-19 0:00:00	U	Si	Juntos	Si	Si	Media	Media	Si	Mayor a	No	Privado	Menor a	2 a 5 hori	No	2009	2	F	2001-04-19 0:00:00	U	No	Separad	Si	Si	Media	Media	Si	Mayor a	No	Privado	Menor a	Menor a	No	2000	2	F	2002-04-19 0:00:00	U	Si	Separad	Si	Si	Secundai	Secundai	No	Mayor a	Si	Privado	Menor a	2 a 5 hori	No	2001	2	M	2000-04-19 0:00:00	U	No	Juntos	Si	Si	Primaria	Media	Si	Mayor a	No	Privado	Menor a	Menor a	No
ano_ingr	periodo	genero	fecha_nacimiento	zona_res	pareja	padres_c	padre_tri	madre_tri	estudios	estudios	vive_con	tamano	discapac	tipo_col	tiempo	v_estudio	prepar																																																																																																																																																																																																																																																																																																																				
2011	1	M	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Media	Si	Menor a	No	Privado	Menor a	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2005	2	F	2000-04-19 0:00:00	U	No	Separad	Si	Si	Profesior	Profesior	Si	Mayor a	Si	Privado	15 a 30 m	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2016	1	M	2002-04-19 0:00:00	U	No	Separad	Si	Si	Media	Secundai	Si	Menor o	Si	Privado	Menor a	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2012	2	F	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Profesior	Profesior	Si	Mayor a	Si	Privado	Menor a	Menor a	No																																																																																																																																																																																																																																																																																																																				
2013	2	F	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Media	No	Mayor a	No	Privado	Menor a	Menor a	No																																																																																																																																																																																																																																																																																																																				
2010	1	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Media	Si	Mayor a	No	Privado	Menor a	Menor a	No																																																																																																																																																																																																																																																																																																																				
2003	2	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Profesior	Profesior	Si	Mayor a	No	Privado	Menor a	2 a 5 hori	Si																																																																																																																																																																																																																																																																																																																				
2006	1	M	2001-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Secundai	Si	Mayor a	No	Privado	Menor a	Menor a	No																																																																																																																																																																																																																																																																																																																				
2008	1	F	2001-04-19 0:00:00	U	No	Separad	Si	Si	Secundai	Secundai	No	Menor o	No	Privado	Menor a	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2009	2	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Profesior	Si	Mayor a	No	Privado	Menor a	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2008	2	M	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Secundai	Media	Si	Mayor a	No	Privado	Menor a	5 a 10 ho	No																																																																																																																																																																																																																																																																																																																				
2013	1	F	2001-04-19 0:00:00	U	Si	Juntos	Si	Si	Secundai	Secundai	Si	Mayor a	No	Privado	Menor a	5 a 10 ho	No																																																																																																																																																																																																																																																																																																																				
2013	2	F	2002-04-19 0:00:00	U	No	Juntos	Si	Si	Media	Secundai	Si	Mayor a	Si	Privado	15 a 30 m	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2015	1	F	2001-04-19 0:00:00	U	Si	Juntos	Si	Si	Media	Media	Si	Mayor a	No	Privado	Menor a	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2009	2	F	2001-04-19 0:00:00	U	No	Separad	Si	Si	Media	Media	Si	Mayor a	No	Privado	Menor a	Menor a	No																																																																																																																																																																																																																																																																																																																				
2000	2	F	2002-04-19 0:00:00	U	Si	Separad	Si	Si	Secundai	Secundai	No	Mayor a	Si	Privado	Menor a	2 a 5 hori	No																																																																																																																																																																																																																																																																																																																				
2001	2	M	2000-04-19 0:00:00	U	No	Juntos	Si	Si	Primaria	Media	Si	Mayor a	No	Privado	Menor a	Menor a	No																																																																																																																																																																																																																																																																																																																				

Figura 67. Resultado de caso de prueba, datos de una hoja de cálculo exportada por la plataforma

Tabla 32

Caso de prueba generar informe sobre deserción

Descripción	Generar informe sobre deserción.
Procedimiento	<ul style="list-style-type: none"> • Iniciar sesión en la plataforma web. • Seleccionar la categoría Deserción. • Seleccionar la opción Informe de deserción.
Resultados esperados	Un documento con información sobre deserción basado en los datos que residen en el sistema.
Resultados obtenidos	La plataforma genera un archivo en formato PDF con la cantidad y porcentaje de posibles desertores, un gráfico de barras, un gráfico de torta y una gráfica de deserción histórica por cohorte.



Resultados

En la fase 1 de la presente investigación se llegó a la conclusión de que la regresión logística fue el modelo de mayor regularidad en cuanto a puntajes obtenidos en medidas de rendimiento realizadas en la fase de evaluación del proceso KDD (matriz de confusión, accuracy, precision, recall y f1-Score), manteniéndose en los primeros puestos en comparación con otros modelos de clasificación (árboles de decisión, vecinos más cercanos, bosque aleatorio y máquina de vectores de soporte), por lo tanto, es el mejor modelo posible para el tratamiento de los datos obtenidos sobre deserción, así mismo presenta una mayor precisión en cuanto a sus predicciones sobre datos nuevos en comparación a los demás modelos, mostrando así una alta tasa de generalización (Pertuz Arroyave & Chaves Sánchez, 2016).

La plataforma web desarrollada en la fase 2 de este proyecto incorpora el modelo de regresión logística seleccionado en la fase 1, en la plataforma web se pueden observar las siguientes medidas de comprobación del modelo luego de su entrenamiento: exactitud o accuracy score, curva de aprendizaje (ver Figura 70), matriz de confusión (ver Figura 71) y un árbol de decisión (ver Figura 72) para interpretar reglas en formato condicional sobre los datos estudiados.

En la Figura 69 se observa una captura de pantalla de la plataforma web mostrando un puntaje de exactitud cercano al 75% obtenido en el entrenamiento del modelo de regresión logística con datos de prueba extraídos de encuestas.

En la curva de aprendizaje se muestra la precisión obtenida tanto con datos de entrenamiento como con datos de prueba en función al número de muestras para detectar con facilidad el grado de dispersión de las predicciones.

Por otra parte, la matriz de confusión permite informar como distribuye los errores un determinado modelo clasificador en una matriz cuadrada donde se suman las predicciones verdaderas positivas (no desertores), verdaderas negativas (desertores), falsas positivas (no desertores erróneos) y falsas negativas (desertores erróneos). La matriz de confusión de la Figura 71 indica una cantidad de 126 predicciones de no desertores correctas, 20 desertores que efectivamente lo son, 10 desertores incorrectos y 39 no desertores incorrectos.

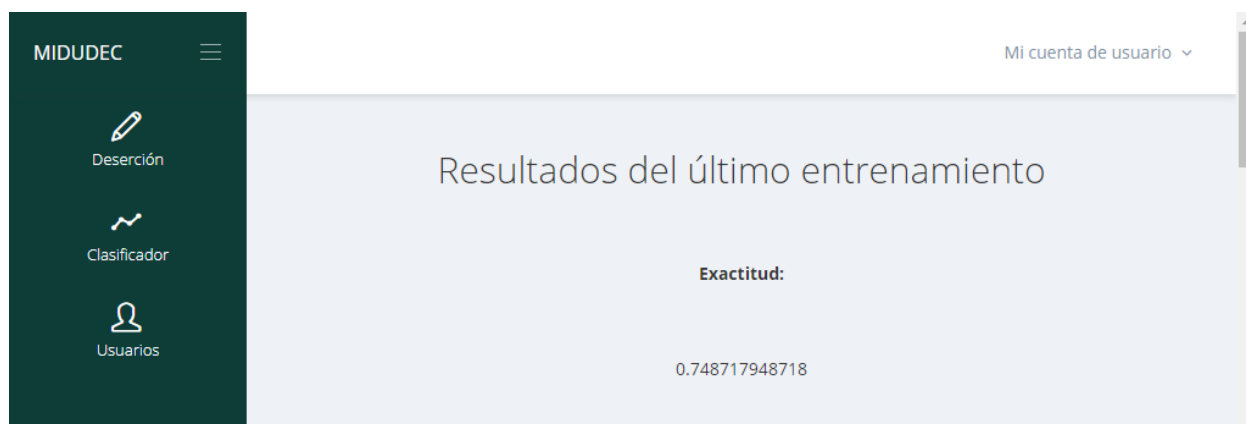


Figura 69. Puntaje accuracy en la plataforma web MIDUDEC

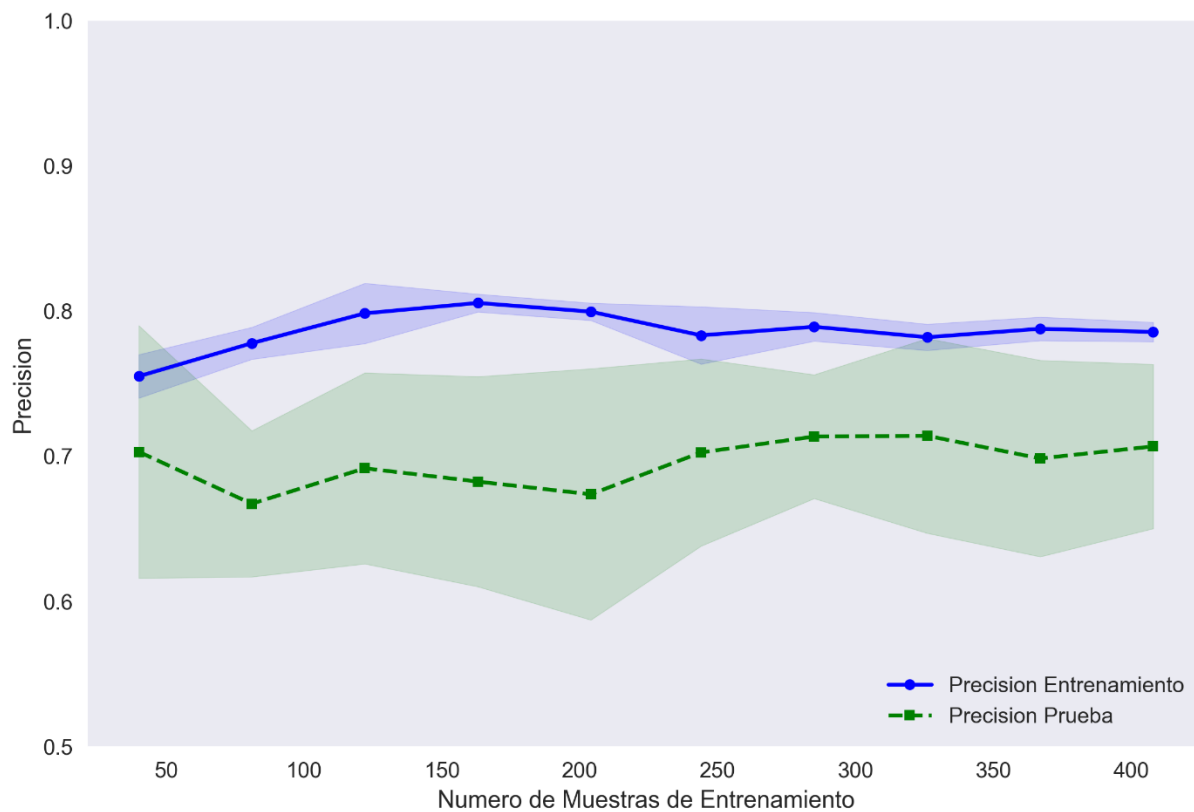


Figura 70. Curva de aprendizaje extraída de la plataforma web MIDUDEC

	0	1
true label 0	126	10
true label 1	39	20
	predicted label	

Figura 71. Matriz de confusión extraída de la plataforma web MIDUDEC

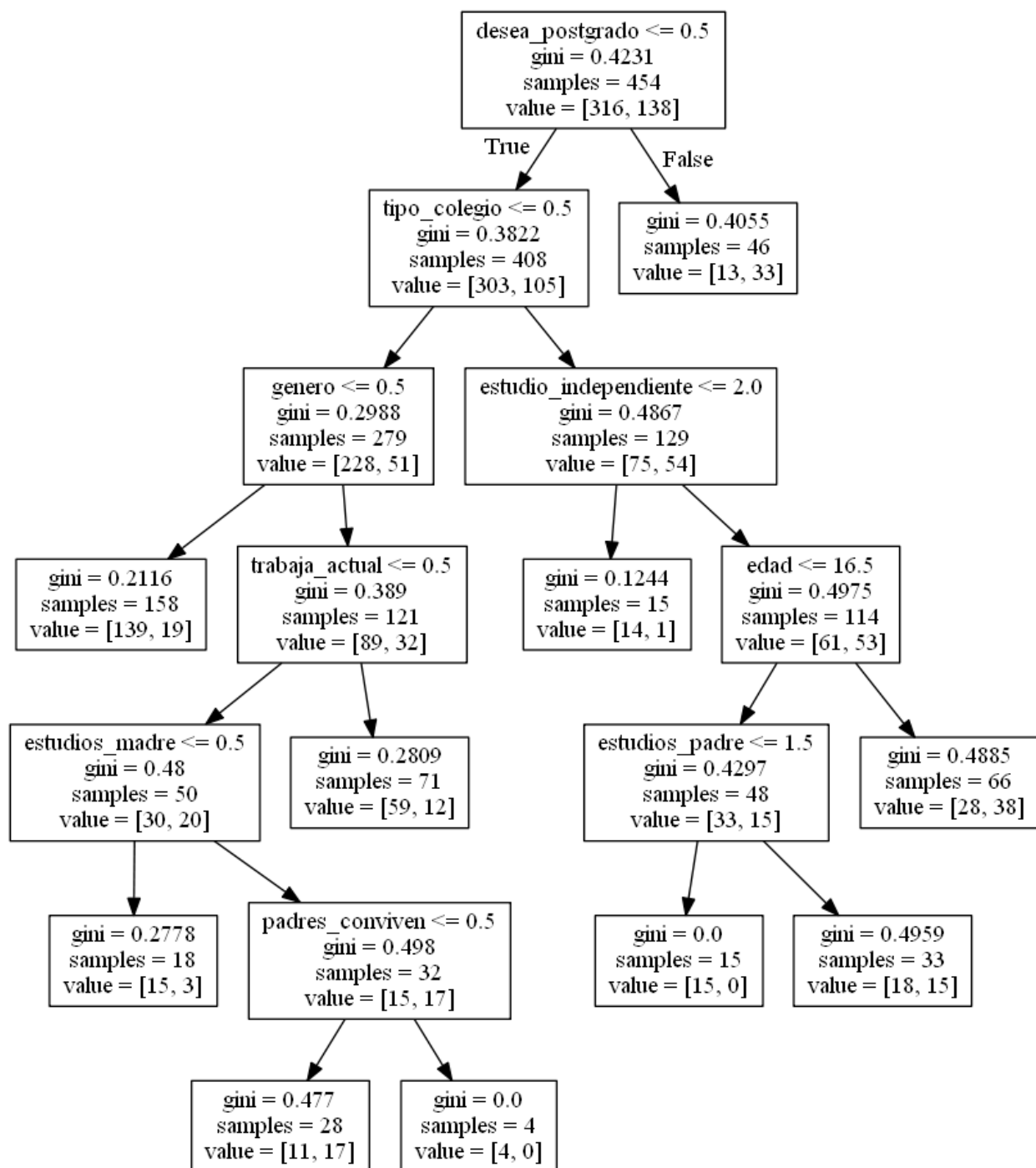


Figura 72. Árbol de decisión extraído de la plataforma web MIDUDEC

Transferencia y divulgación del conocimiento

Se elaboró un artículo científico para participar en el XII Congreso Internacional de Ingeniería "Electrónica y Tecnologías Avanzadas" - CIETA. VIII Congreso Internacional en Sistemas, Informática e Ingeniería del Conocimiento - CICOM. VIII Congreso Internacional en Telecomunicaciones, el VII Congreso Internacional en Ingeniería Eléctrica y el III Congreso Internacional de Ingeniería Mecatrónica, que se celebró en la ciudad de Pamplona los días 10, 11 y 12 de mayo de 2017.

El día 10 de mayo de 2017 se realizó una ponencia en el XII Congreso Internacional Electrónica y Tecnologías de Avanzada en la Universidad de Pamplona, ubicada en la ciudad de Pamplona, Norte de Santander, Colombia.

También se elaboró un artículo científico con la intención de participar en el 1er. Congreso Latinoamericano de Ingeniería (CLADI 2017) que se realizó en Paraná y Oro Verde, Entre Ríos, Argentina del 13 al 15 de septiembre de 2017.

Se realizó además un artículo para participar en el I Congreso Internacional de Tecnología, Ingeniería e Innovación (CITII - 2017) que se realizó en las instalaciones de la Universidad de San Buenaventura, sede Bogotá, (Colombia) del 18 de octubre al 20 de octubre de 2017.

Se elaboró un artículo y un poster para participar en el Primer Encuentro Académico de Creatividad e Innovación realizado en la Universidad de Cundinamarca, extensión Facatativá los días 8 y 9 de septiembre de 2017. En el evento mencionado se realizó la exposición del poster y la presentación de una ponencia basada en el artículo.

Por último, se presentó una ponencia el día 8 de noviembre de 2017 en el VI Congreso Internacional de Ingeniería de la Universidad de Cundinamarca – CIIUDEC 2017.

Conclusiones y recomendaciones

La Universidad de Cundinamarca no cuenta con una consolidación de perfiles característicos de cada estudiante que permitan estudiar el problema de la deserción con el enfoque de la presente investigación.

La plataforma web permite analizar el problema de la deserción en la Universidad de Cundinamarca de forma más específica en torno a las variables influyentes que con datos porcentuales como se realiza actualmente.

Las variables a analizar para predecir la deserción son un estándar determinado por el MEN, aunque ciertas variaciones a ese estándar muestran que existen otras variables de tipo personal que pueden afectar el proceso en gran medida y que requieren ser estudiadas de igual forma.

El proceso KDD hace posible indagar y encontrar iterativamente un modelo adecuado para un dominio de estudio. En la presente investigación se evidencia el modelo de regresión logística como el mejor método encontrado para los datos obtenidos, debido a su mayor regularidad en cuanto a puntajes logrados en las distintas pruebas a diferencia del resto de modelos evaluados.

Se sugiere realizar gestiones para conseguir datos históricos de la Universidad de Cundinamarca, que permitirán obtener resultados en las predicciones más certeros y cercanos a la realidad de la institución que los obtenidos en esta investigación.

Se recomienda realizar nuevas iteraciones del proceso KDD para evaluar más modelos de clasificación. La interfaz RESTful facilita realizar cambios de modelo predictivo sin tener que modificar otros componentes de la plataforma web.

Referencias

- Amaya, Y., Barrientos, E., & Santander, P. (2014). Modelo predictivo de deserción estudiantil.
- Azoumana, K. (2013). Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos., *6*(10), 41–51.
- Bravo Castillo, M., & Mejía Giraldo, A. (2010). Los retos de la educación superior en Colombia: Una reflexión sobre el fenómeno de la deserción universitaria. *Revista Educación En Ingeniería*, *5*(10), 85–98. Recuperado de <http://www.educacioneningeneria.org/index.php/edi/article/view/101>
- Colombia, M. de comercio industria y turismo de. (2013). Decreto 1377 de 2012. *Ministerio de Comercio, Industria Y Turismo, decreto 13*, 1–11. Recuperado de http://www.mintic.gov.co/portal/604/articulos-4274_documento.pdf
- Colombia, M. de E. de. (2015). Estadísticas Deserción y Graduación 2015, 4.
- Congreso de la República de Colombia. (1982). LEY NÚMERO 23 DE 1982 "Sobre derechos de autor" *Congreso de La República de Colombia, 1982*, 57. Recuperado de http://www.uptc.edu.co/export/sites/default/editorial/documentos/Ley_23_1982.pdf
- Congreso de la República de Colombia. (2009). Ley 1273, (48).
- Congreso de la República de Colombia. (2015). Ley estatutaria 1757.
- Corte constitucional. (1991). Constitución Política de Colombia. Recuperado de <http://www.corteconstitucional.gov.co/inicio/Constitucion politica de Colombia - 2015.pdf>
- Emanuel Chengu. (2007). Manual de consulta del usuario SPADIES version 2.8, 82.
- Fundación Universitaria Luis Amigó. (2013). Análisis general de deserción estudiantil.
- Grinberg, M. (2015). *Flask Web Development*. O'REILLY.
- Guzmán Ruiz, C., Muriel Durán, D., & Franco Gallego, J. (2009). *Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para su prevención*. Recuperado de http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articulos-254702_libro_desercion.pdf
- Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2004). *Introducción a la Minería de Datos*. Prentice Hall.
- Klein, B. (2015). History of Python. *Python Course*, 1–3. Recuperado de http://www.python-course.eu/python3_history_and_philosophy.php
- Kuna Horacio, García Martínez Ramón, V. F. R. (2010). Pattern Discovery In University Students Desertion Based On Data Mining. *Proceedings of The IV Meeting on Dynamics of Social and Economic Systems*, *2*(2), 11. Recuperado de <http://idia.com.ar/rgm/articulos/AASSJ-4091023-SD48-Kuna.pdf>
- Luan, J. (2002). Data Mining and Knowledge Management in Higher Education. *Workshop*

Associate of Institutional Research International Conference, Toronto, 1–18.

- Lucchi, R., Millot, M., & Elfers, C. (2008). Resource Oriented Architecture and REST. *Assessment of Impact and Advantages on INSPIRE, Ispra: European Communities, 5–13.* <https://doi.org/10.2788/80035>
- Marcano, Y. J., & Rodríguez, R. H. (2014). Licenciatura En Computación-Universidad Del Zulia-Npf Data Mining Applied To School Dropout Case : Bachelor of Computing At Universidad Del Zulia-Npf, 31–51.
- Márquez Vera, C. (2015). Predicción del fracaso y abandono escolar mediante técnica de minería de datos, 1–133. Recuperado de <http://helvia.uco.es/xmlui/bitstream/handle/10396/12852/2015000001157.pdf?sequence=1>
- Martinez, M. G. Y. F. (2004). Confrontacion de dos tecnicas de Minería de datos aplicados a un dominio especifico. *Pontifica, 125.* Recuperado de <http://www.javeriana.edu.co/biblos/tesis/ingenieria/Tesis184.pdf>
- Méndez, N. D. J. C. (2006). Minería De Datos Una Herramienta Para La Toma De Decisiones.
- Merchán, S. M., & Duarte, J. A. (2016). Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance, *14(6), 2783–2788.* <https://doi.org/10.1109/ICEEE.2007.4344994>
- Moine, J. M. (2013). Metodologías para el descubrimiento de conocimiento en bases de datos : un estudio comparativo, 111.
- 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.*
- Morales, S. M. (2012). Uso de la Metodología Rough Sets para la identificación de variables críticas influyentes en una base de datos. Recuperado de <http://intellectum.unisabana.edu.co:8080/jspui/handle/10818/1459>
- Ordóñez, K. (2013). Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes de la modalidad abierta y a distancia, 260. Recuperado de [http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez Briceño Karla-Infomatica.pdf](http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez%20Brice%C3%B1o%20Karla-Infomatica.pdf)
- Orea, S. V. (2009). Aplicación de técnicas de minería de datos para predecir deserción. ... *Técnicas De Minería De Datos Para Predecir Deserción.* Recuperado de <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:APLICACION+DE+TECNICAS+DE+MINERIA+DE+DATOS+PARA+PREDECIR+DESERCIÓN>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python, *12, 2825–2830.* <https://doi.org/10.1007/s13398-014-0173-7.2>
- Pereira, R. T. (2009). Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos. *Iiis.org.* Recuperado de <http://www.iiis.org/CDs2009/CD2009CSC/CISCI2009/PapersPdf/C692YV.pdf>

- Pertuz Arroyave, C. J., & Chaves Sánchez, L. J. (2016). Aplicación de técnicas de Machine Learning para el proceso de minería de datos en la identificación de patrones y perfiles característicos asociados a la deserción universitaria.
- Pinzon Cadena, L. L. (2011). Aplicando minería de datos al marketing educativo. *Notas D Marketing*, 1(1), 45–61. Recuperado de <http://www.usergioarboleda.edu.co/investigacion-marketing/marketing/articulo5MineriaDatos.pdf>
- Reid, M. D., Montgomery, J., & Drake, B. (2016). Protocols and Structures for Inference : A RESTful API for Machine Learning. *PAPIs 2015*, 29–42. Recuperado de <http://jmlr.org/proceedings/papers/v50/montgomery15.pdf>
- Reinoso, G. B. (1999). Modelado Orientado a Objetos : Una Evaluación Crítica. *Revista Facultad de Ingeniería*.
- Ribeiro Lima, I., Castro Freire, T. de, & Costa, H. A. X. (2012). Adapting and Using Scrum in a Software Research and Development Laboratory. *Revista de Sistemas de Informação Da FSMA*, 9, 16–23. Recuperado de http://www.fsma.edu.br/si/edicao9/FSMA_SI_2012_1_Principal_2_en.pdf
- Richardson, L., & Ruby, S. (2007). *RESTful Web Services*. Vasa. <https://doi.org/10.1109/MIC.2008.130>
- Rodriguez, A. (2008). Restful web services: The basics. *Online Article in IBM DeveloperWorks Technical Library*, (November), 1–11. Recuperado de <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:RESTful+Web+services+:+The+basics#0>
- Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto, S. (2012). Predicción del Fracaso Escolar Mediante Técnicas de Minería de Datos. *Iee-Rita*, 7(3), 109–117.
- Sah, N. K. (2013). Thesis Degree Programme in Chemistry & Technology, (April).
- Sautu, R. (2005). Manual de metodología.
- Schwaber, K., & Sutherland, J. (2013). The Scrum Guide. *Scrum.Org and ScrumInc*, (July), 17. <https://doi.org/10.1053/j.jrn.2009.08.012>
- Sposito, O., & Etcheverry, M. (2010). Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil Deserción. ... *En Sistemas, Cibernética E ...*. Recuperado de /citations?view_op=view_citation&continue=/scholar?hl=es&start=20&as_sdt=0,5&scilib=1024&citilm=1&citation_for_view=uOPNWhoAAAAJ:hNSvKAmkeYkC&hl=es&oi=p
- Universidad de Cundinamarca. (2015). Boletín estadístico VII.
- Universidad del los Andes. (2002). Determinantes de la Deserción. “*Informe Mensual Sobre El Soporte Técnico Y Avance Del Contrato Para Garantizar La Alimentación, Consolidación, Validación Y Uso de La Información Del SPADIES*,” 25(49), 122–164. Recuperado de http://revistas.pucp.edu.pe/index.php/economia/article/viewFile/944/911%5Cnhttp://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702_Informe_determinantes_desercion.pdf

- Veitch, W. R. (2010). Identifying Characteristics of High School Dropouts: Data Mining With a Decision Tree Model. *EricWeb*, 1–11. Recuperado de http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED490086&ERICExtSearch_SearchType_0=no&accno=ED490086
- Virveda Benito, F., & Román Carrillo, J. (2010). Minería de datos y aplicaciones. *Revista Latinoamericana de Ingeniería de Software*, 8. Recuperado de <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>

Anexos

En el CD institucional se encuentran los siguientes documentos que fueron de vital importancia para la investigación y ejecución del proyecto:

- Manual de usuario: Instalación, configuración y uso de la plataforma web MIDUDEC
- Artículos científicos: Artículos realizados para la participación en ponencias de congresos nacionales e internacionales.
- Dataset: Un archivo de hoja de cálculo que contiene los perfiles de estudiantes usados en la investigación.
- Código: Carpeta que contiene el código fuente de la interfaz RESTful y de la plataforma web MIDUDEC.
- MINr008: Documento preliminar que describe la necesidad del problema y la propuesta de solución.