

	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA			
	<small>Documento</small> FORMATO HOJA DE RESUMEN PARA TRABAJO DE GRADO	<small>Código</small> F-AC-DBL-007	<small>Fecha</small> 10-04-2012	<small>Revisión</small> A
	<small>Dependencia</small> DIVISIÓN DE BIBLIOTECA	<small>Aprobado</small> SUBDIRECTOR ACADEMICO		<small>Pág.</small> i(166)

RESUMEN – TRABAJO DE GRADO

AUTORES	BRENDA YURIMAR CHINDOY CHASOY KELY YINETH DIAZ PEDROZA		
FACULTAD	INGENIERÍA		
PLAN DE ESTUDIOS	INGENIERÍA DE SISTEMAS		
DIRECTOR	ALVEIRO ALONSO ROSADO GOMEZ		
TÍTULO DE LA TESIS	DESARROLLO DE UN SOFTWARE QUE PERMITA LA DETECCIÓN DE ESTUDIANTES DESERTORES DEL PROGRAMA DE INGENIERÍA DE SISTEMAS UTILIZANDO PROCESOS Y ELEMENTOS DE MINERÍA DE DATOS		
RESUMEN (70 palabras aproximadamente)			
<p>DESARROLLO DE UN SOFTWARE QUE PERMITA LA DETECCIÓN DE ESTUDIANTES DESERTORES DEL PROGRAMA DE INGENIERÍA DE SISTEMAS UTILIZANDO PROCESOS Y ELEMENTOS DE MINERÍA DE DATOS, PARTE DE UNA REVISIÓN LITERARIA ENTORNO A LA DESERCIÓN Y EL TRATAMIENTO DE ESTA POR MEDIO DE MINERÍA DE DATOS, PASANDO POR PRÁCTICAS MANUALES PARA LA PREDICCIÓN DEL RENDIMIENTO ACADÉMICO, DE LO CUAL SE DETECTÓ QUE ES EL PRINCIPAL CAUSANTE Y EL MAYOR FOCO PARA ATACAR Y DISMINUIR LOS ÍNDICES DE DESERCIÓN DENTRO DE LA CARRERA.</p> <p>COMO RESULTADO DE ESTA INVESTIGACIÓN SE DESEA PROPONER: UN SOFTWARE QUE AUTOMATICE EL PROCESO DE LIMPIEZA DE DATOS Y QUE ADEMÁS CONTENGA LA CREACIÓN DEL MODELO Y LA PREDICCIÓN, CON EL FIN DE PLANTEAR ESTRATEGIAS PREVIAS AL DESARROLLO DEL SEMESTRE, DE LOS ESTUDIANTES QUE SE INGRESE EN EL SISTEMA.</p>			
CARACTERÍSTICAS			
PÁGINAS: 166	PLANOS:	ILUSTRACIONES: 53	CD-ROM:1



**DESARROLLO DE UN SOFTWARE QUE PERMITA LA DETECCIÓN DE
ESTUDIANTES DESERTORES DEL PROGRAMA DE INGENIERÍA DE SISTEMAS
UTILIZANDO PROCESOS Y ELEMENTOS DE MINERÍA DE DATOS**

AUTORES:

KELY YINETH DIAZ PEDROZA

BRENDA YURIMAR CHINDOY CHASOY

Proyecto realizado como requisito para optar por el título de ingeniero de sistemas

Director

ALVEIRO ALONSO ROSADO GOMEZ

Msc. Gestión, aplicación y desarrollo software

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA

FACULTAD DE INGENIERIAS

INGENIERÍA DE SISTEMAS

Ocaña, Colombia

octubre, 2019

Dedicatoria

La presente tesis está dedicada primeramente a Dios, ya que gracias a él logramos llegar hasta este momento tan importante de nuestra formación profesional.

A nuestras madres, Fany Chasoy e Irene Pedroza, por ser nuestra inspiración en todo momento, el pilar más importante demostrándonos siempre su cariño y apoyo incondicional, nuestros padres que están en el cielo que se encargaron de guiarnos y cuidarnos en todo el proceso, y que los hechamos mucho de menos.

¡GRACIAS A TODOS!

Agradecimientos

Al creador de todas las cosas, quien nos dio la fuerza y fortaleza para continuar cuando estábamos a punto de desfallecer, con toda la humildad de nuestro corazón, dedicamos primeramente nuestro trabajo a Dios.

Agradecemos enormemente a todas las personas que han ayudado y apoyado a lo largo de estos años de carrera profesional.

El amor recibido, la dedicación y la paciencia con la que cada día se preocupaban nuestras madres, por nuestro avance y desarrollo de esta tesis, es simplemente único y se refleja en nosotras, sus hijas.

Gracias a nuestras madres por ser las principales promotoras de nuestros sueños, gracias a ellas por cada día confiar y creer en nosotras y en nuestras expectativas, gracias a ellas por estar dispuestas a acompañarnos cada larga y agotadora noche de estudio, gracias por cada consejo y por cada una de las palabras que nos guiaron durante nuestra vida.

A nuestro director de tesis, Alveiro Alonso Rosado Gómez quien fue nuestro guía en todo el proceso y al cual apreciamos por su gran calidad humana; sin duda alguna, su ayuda fue incondicional, compartiendo su conocimiento y experiencia para culminar esta etapa tan importante de nuestras vidas.

Finalmente a todas las personas que contribuyeron y aportaron su granito de arena para la elaboración de este proyecto.

Índice

Capítulo 1. Desarrollo de un software que permita la detección de estudiantes desertores del programa de ingeniería de sistemas utilizando procesos y elementos de minería de datos	1
1.1 Planteamiento del problema	1
1.2 Formulación del problema	3
1.3 Objetivos	3
1.3.1 Objetivo general	3
1.3.2 Objetivos específicos	3
1.4 Justificación	4
1.5 Delimitaciones	5
1.5.1 Delimitación operativa	5
1.5.2 Delimitación conceptual	6
1.5.3 Delimitación geográfica	6
1.5.4 Delimitación temporal	6
Capítulo 2. Marco referencial	7
2.1 Marco histórico	7
2.1.1 Revisión de la deserción estudiantil a nivel internacional	7
2.1.2 Revisión de la deserción estudiantil a nivel nacional	9
2.1.3 Revisión de la deserción estudiantil a nivel regional	14
2.2 Marco contextual	15
2.3 Marco conceptual	15
2.3.1 Deserción estudiantil	15
2.3.2 Desertor	16
2.3.3 Java	16
2.3.4 Descubrimiento de conocimiento en bases de datos (KDD)	16
2.3.5 Minería de datos	16
2.3.6 Patrones	17
2.3.7 Permanencia estudiantil	17
2.3.8 Sistema de información SPADIES	17
2.3.9 Software	18
2.3.10 Machine learning	18

2.3.11	Weka.....	18
2.4	Marco teórico.....	18
2.4.1	Teorías de la deserción estudiantil.	18
2.4.2	Tipos de deserción.....	22
2.4.3	Teorías la minería de datos en la deserción estudiantil.	23
2.5	Marco legal.....	24
2.5.1	Capítulo I de la ley general de educación.....	24
2.5.2	Artículo 15 de la constitución política de Colombia.....	24
2.5.3	La ley 30 de 1992 en su título V, capítulo III, artículo 117.....	24
2.5.4	Ley general de educación en su artículo 4 establece.....	25
2.5.5	Ley 1188 de 2008.....	25
2.5.6	Ley 1273 de 2009.....	25
2.5.7	Artículo 76 de la ley No. 112 de 2011.....	25
2.5.8	La ley estatutaria 1581 de 2012.....	26
2.5.9	Ley 1620 de 2013.....	26
2.5.10	Acuerdo nacional para disminuir la deserción en educación superior.....	26
2.5.11	Ley 1712 de 2014.....	26
2.5.12	CONPES 3920 de 2018.....	27
Capítulo 3.	Diseño metodológico	28
3.1	Tipo de investigación.....	28
3.2	Metodología de la investigación.....	28
3.3	Población.....	29
3.4	Muestra.....	29
3.5	Técnicas e instrumentos de recolección.....	29
3.6	Análisis de la información.....	29
Capítulo 4.	Resultados	31
4.1	Conocimiento del entorno en el que se pretende desarrollar el proyecto	31
4.2	Descubrimiento de conocimiento en bases de datos (KDD).....	31
4.2.1	Selección de datos.....	32
4.2.2	Pre-procesamiento.....	36
4.2.3	Transformación.....	39

4.2.4	Minería de datos.	44
4.2.5	Evaluación e interpretación.	59
4.3	Desarrollo del software.	60
4.3.1	Roles.	61
4.3.2	Fases.	61
4.4.	Análisis e interpretación de los resultados obtenidos.	125
Capítulo 5. Conclusiones.		126
Capítulo 6. Recomendaciones.		127
Referencias.		130
Apéndices.		136

Lista de tablas

Tabla 1	<i>Atributos seleccionados según la literatura</i>	32
Tabla 2	<i>Datos sistemas</i>	35
Tabla 3	<i>Atributos eliminados</i>	37
Tabla 4	<i>Discretización del atributo sanción</i>	39
Tabla 5	<i>Discretización del atributo año de ingreso</i>	39
Tabla 6	<i>Discretización del atributo semestre actual</i>	40
Tabla 7	<i>Discretización del atributo valoración icfes</i>	40
Tabla 8	<i>Discretización del atributo edad</i>	41
Tabla 9	<i>Discretización del atributo estrato social</i>	42
Tabla 10	<i>Discretización del atributo lugar de procedencia</i>	42
Tabla 11	<i>Discretización del atributo departamento de procedencia</i>	42
Tabla 12	<i>Discretización del atributo número de hermanos estudiando</i>	43
Tabla 13	<i>Discretización del atributo estado civil</i>	43
Tabla 14	<i>Discretización del atributo promedio acumulado</i>	44
Tabla 15	<i>Selección de atributos con WEKA</i>	45
Tabla 16	<i>Selección de atributos con métodos y algoritmos de evaluación (ordenamiento y ranqueo)</i>	46
Tabla 17	<i>J48 con factor de confianza</i>	47
Tabla 18	<i>Clasificación de instancias Naive Bayes</i>	50
Tabla 19	<i>Clasificación de instancias con RandomizableFilteredClassifier</i>	53
Tabla 20	<i>Evaluación simple</i>	56
Tabla 21	<i>Evaluación cruzada</i>	58
Tabla 22	<i>Validación Split</i>	59
Tabla 23	<i>Predicción</i>	59
Tabla 24	<i>Roles</i>	61
Tabla 25	<i>Registrar usuario</i>	67
Tabla 26	<i>Loguear usuario</i>	68
Tabla 27	<i>Gestionar principal</i>	69
Tabla 28	<i>Gestionar proceso</i>	70
Tabla 29	<i>Crear proceso</i>	71
Tabla 30	<i>Consultar proceso</i>	71
Tabla 31	<i>Cargar datos entrenamiento</i>	72
Tabla 32	<i>Consultar la información de los datos de entrenamiento</i>	73
Tabla 33	<i>Convertir los datos entrenamiento al formato arff</i>	74
Tabla 34	<i>Crear modelo</i>	75
Tabla 35	<i>Guardar modelo</i>	76
Tabla 36	<i>Consultar modelo</i>	77

Tabla 37	<i>Cargar datos de prueba</i>	78
Tabla 38	<i>Consultar la información de los datos de prueba</i>	79
Tabla 39	<i>Convertir los datos de prueba al formato arff</i>	80
Tabla 40	<i>Crear predicción</i>	81
Tabla 41	<i>Guardar predicción</i>	82
Tabla 42	<i>Consultar predicción</i>	83
Tabla 43	<i>Pruebas de integración</i>	119
Tabla 44	<i>Líneas de código: paquetes Vistas</i>	120
Tabla 45	<i>Líneas de código: paquete Controlador</i>	121
Tabla 46	<i>Líneas de código: paquete experto</i>	121
Tabla 47	<i>Líneas de código: paquete fachada</i>	121
Tabla 48	<i>Líneas de código: paquete modelo</i>	122
Tabla 49	<i>Líneas de código: paquete util</i>	122
Tabla 50	<i>Líneas de código: paquete principal</i>	123
Tabla 51	<i>Total de líneas de código software</i>	123

Lista de figuras

<i>Figura 1.</i> Matriz de confusión J48.....	48
<i>Figura 2.</i> Matriz de confusión Naive Bayes	51
<i>Figura 3.</i> Matriz de confusión RandomizableFilteredClasiffier	54
<i>Figura 4.</i> Diagrama Modelo-Vista-Controlador.	64
<i>Figura 5.</i> Diagrama de actores.	65
<i>Figura 6.</i> Diagrama de casos de uso.	66
<i>Figura 7.</i> Registrar usuario.	84
<i>Figura 8.</i> Loguear usuario.	85
<i>Figura 9.</i> Gestionar principal.	86
<i>Figura 10.</i> Gestionar proceso.	87
<i>Figura 11.</i> Gestionar información del proyecto: Crear proceso.	88
<i>Figura 12.</i> Gestionar información del proyecto: Consultar proyecto.	89
<i>Figura 13.</i> Gestionar datos entrenamiento: Cargar datos de entrenamiento.	90
<i>Figura 14.</i> Gestionar datos entrenamiento: Consultar datos de entrenamiento.	91
<i>Figura 15.</i> Gestionar minería: Conversión de los datos de entrenamiento al formato arff.	92
<i>Figura 16.</i> Gestionar minería: Crear modelo.	93
<i>Figura 17.</i> Gestionar minería: Guardar modelo.	94
<i>Figura 18.</i> Gestionar minería: Consultar modelo.	94
<i>Figura 19.</i> Gestionar datos prueba: Cargar datos de prueba.	95
<i>Figura 20.</i> Gestionar datos prueba: Consultar datos de prueba.	96
<i>Figura 21.</i> Gestionar predicción: Conversión de los datos de prueba al formato arff.	97
<i>Figura 22.</i> Gestionar predicción: Crear predicción.	98
<i>Figura 23.</i> Gestionar predicción: Guardar predicción.	99
<i>Figura 24.</i> Gestionar predicción: Consultar predicción.	99
<i>Figura 25.</i> Modelo relacional.	100
<i>Figura 26.</i> Arquitectura por capas.	101
<i>Figura 27.</i> Diagrama de paquetes.	102
<i>Figura 28.</i> Diagrama de clase: controlador.	103
<i>Figura 29.</i> Diagrama de clase: experto.	104
<i>Figura 30.</i> Diagrama de clase: fachada.	105
<i>Figura 31.</i> Diagrama de clase: conexión.	106
<i>Figura 32.</i> Diagrama de clase: dao.	107
<i>Figura 33.</i> Diagrama de clase: dto.	108
<i>Figura 34.</i> Diagrama de clase: estrategia.	109
<i>Figura 35.</i> Diagrama de clase: útil.	110

<i>Figura 36.</i> Diagrama de clase: vista.	111
<i>Figura 37.</i> Diagrama de secuencia: Usuario.	112
<i>Figura 38.</i> Diagrama de secuencia: Login.	112
<i>Figura 39.</i> Diagrama de secuencia: Agregar proceso.	113
<i>Figura 40.</i> Diagrama de secuencia: Arff entrenamiento.	113
<i>Figura 41.</i> Diagrama de secuencia: Arff prueba.	114
<i>Figura 42.</i> Diagrama de secuencia: Cargar datos de entrenamiento.	114
<i>Figura 43.</i> Diagrama de secuencia: Cargar datos de prueba.	115
<i>Figura 44.</i> Diagrama de secuencia: Clave maestra.	115
<i>Figura 45.</i> Diagrama de secuencia: Guardar modelo.	115
<i>Figura 46.</i> Diagrama de secuencia: Consultar modelo.	116
<i>Figura 47.</i> Diagrama de secuencia: Consultar predicción.	116
<i>Figura 48.</i> Diagrama de secuencia: Guardar predicción.	117
<i>Figura 49.</i> Diagrama de secuencia: Modelo.	117
<i>Figura 50.</i> Diagrama de secuencia: Mostrar datos de entrenamiento.	118
<i>Figura 51.</i> Diagrama de secuencia: Mostrar datos de prueba.	118
<i>Figura 52.</i> Diagrama de secuencia: Predicción.	118
<i>Figura 53.</i> Diagrama de secuencia: Proceso.	119

Capítulo 1. Desarrollo de un software que permita la detección de estudiantes desertores del programa de ingeniería de sistemas utilizando procesos y elementos de minería de datos

1.1 Planteamiento del problema

Hoy en día el estado colombiano, organizaciones e instituciones tanto públicas como privadas han realizado varios estudios acerca de todo lo que conlleva la deserción estudiantil, ya que es una de las principales problemáticas que aqueja a las instituciones de educación superior (IES) en el país, generada por diferentes causas que son difíciles de identificar a simple vista, entre ellas están las académicas, económicas, sociales, psicológicas, entre otras (Universidad de los Andes, 2014). Uno de los estudios es el que se realizó por el banco mundial, en el que se obtuvo una estadística que evidencia que para el 2017 Colombia ocupaba el segundo puesto en deserción estudiantil luego de Bolivia con una aproximación casi igual al porcentaje de graduados en educación superior en América Latina y el Caribe (Ferreyra, Avitabile, Botero, Haimovich, & Urzúa, 2017), dando una alerta a las IES, para que estas con el acompañamiento del estado, generen nuevas medidas de prevención ante esta la deserción basadas en estudios detallados de ingeniería. Este problema genera un alto costo que compromete la estabilidad de los ingresos transferidos a las Instituciones de educación superior, la calidad de los procesos administrativos y académicos. Así mismo, se constituye en un alto costo social el cual es asumido por las familias, la universidad y el Estado (Rueda & Socha, 2009).

Hay que tener claro que para el ministerio de educación nacional (MEN), un desertor es aquel que presenta inactividad académica durante dos semestres académicos consecutivos

(MEN, 2012) de tal manera que dicha definición trasciende a tres tipos de abandono estudiantil en el pregrado, tales como: deserción precoz cuando un estudiante es admitido y no se matricula, temprana cuando abandonan los estudios en los primeros semestres, estos dos, presentan el porcentaje más alto; y por último la deserción tardía cuando abandonan sus estudios en los últimos semestres siendo esta la más grave, cada una de estas situaciones se dan por diferentes causas, pues son periodos en las que pueden ir cambiando ciertas variables como las materias que estos estén cursando o el nivel de madurez económica que hayan adquirido (Guzmán et al., 2009), convirtiéndose en el punto clave para tratar este fenómeno.

La universidad Francisco de Paula Santander seccional Ocaña cuenta con el programa de ingeniería de sistemas desde hace más de 6 años, en donde presenta un problema real, en años anteriores al 2017 se presentaba entre el 40% y 70% de deserción estudiantil, convirtiéndose en la segunda carrera dentro de la facultad de ingenierías con mayor deserción; las causas del abandono no se tienen del todo claras, pues, no se han realizado estudios profundos en cada caso de deserción y no se sabe por qué los estudiantes tomaban la decisión de desertar, haciendo difícil el camino hacia la meta que tiene el MEN de reducir el índice de deserción a un 9% del porcentaje anual en las IES. Las estrategias que se han realizado no cuentan con un análisis minucioso, es por ello que se ve necesario realizar un estudio que permita abordar el tema de deserción estudiantil en el caso preciso del programa de ingeniería de sistemas bajo una herramienta que facilite el análisis y un seguimiento constante de las diferentes alertas de abandono por parte de los estudiantes.

1.2 Formulación del problema.

¿La aplicación de técnicas de minería de datos permitirá la construcción de un software que pueda predecir la probabilidad de desertar que tiene un estudiante del programa de ingeniería de sistemas de la universidad Francisco de Paula Santander Ocaña?

1.3 Objetivos

1.3.1 **Objetivo general.** Desarrollar un software que permita la detección de estudiantes desertores del programa de ingeniería de sistemas utilizando procesos y elementos de minería de datos.

1.3.2 **Objetivos específicos.** Caracterizar las variables relevantes en la deserción con base a una revisión de la literatura sobre este tema.

- Aplicar el proceso de extracción del conocimiento en base datos con el fin de establecer el conjunto de datos con los que se trabajara y la técnica de minería de datos que mejor se ajuste a ellos.

- Construir una herramienta informática que prediga la probabilidad que tiene un estudiante de ser desertor.

- Analizar los resultados obtenidos a través de la herramienta de deserción estudiantil.

1.4 Justificación

Para el MEN existen diferentes métodos para la adaptación y transferencia, una de ellas es la que se mencionara a continuación ya que es una de las medidas que las instituciones deberían tener.

Programas o sistemas de alertas tempranas para identificar el riesgo de deserción por estudiante. La herramienta debe dirigirse a mejorar la caracterización estudiantil, a integrar el seguimiento académico, la información de SPADIES y demás variables del entorno institucional que garanticen el tratamiento integral y preventivo del abandono de estudios. (Acuerdo nacional para disminuir la deserción en educación superior, 2013).

Frente a esta situación de deserción, el plan de estudios de ingeniería de sistemas tomó una decisión radical acorde al análisis manual que realizaron; reducir el número de estudiantes que pudiesen ser admitidos por segunda opción, permitiendo aceptar a estudiantes que realmente deseen estudiar ingeniería de sistemas, resultando un tanto favorable ya que el programa ha reducido el nivel de deserción estudiantil. Sin embargo, el plan de estudios manifiesta que una de las causas de deserción estudiantil es la desorganización de la información y no se ha realizado un análisis profundo que permita obtener información valiosa de aquellos datos almacenados y así saber realmente por qué ocurre el abandono por parte de los estudiantes, ya sea por la falta de recursos económicos, problemas sociales, familiar o dificultades académicas.

Por lo anterior, se pretende contribuir a mejorar el plan de estudio generando una herramienta que permita predecir las variables de mayor incidencia en la deserción y reducir el porcentaje de abandono, escudriñando esos datos del estudiantado utilizando minería de datos y con esto poder “extraer conocimiento útil, comprensible grandes volúmenes de datos, donde inicialmente ese conocimiento es desconocido pero que se encuentra en las relaciones implícitas de los atributos” (Ordoñez Leyva, Avílez Vásquez, & Grass Boada, 2011).

La herramienta busca encontrar características en los estudiantes que puedan ser capaces de provocar la deserción, en esta estarán presentes técnicas de minería de datos y los atributos previamente establecidos dentro de la base de datos que se pretende usar para hacer una detección de los factores por los que desertan estos estudiantes.

1.5 Delimitaciones

1.5.1 Delimitación operativa. La creación de la herramienta se hará con base a los objetivos planteados, en este proceso es posible que se tengan inconvenientes con el suministro de los datos por parte de las fuentes necesarias para la investigación debido a la confidencialidad de los datos.

- Realizar una búsqueda acerca de los estudios realizados por diferentes entidades, el estado, organizaciones e instituciones para la identificación de variables que mayor influyen en la deserción.
- Creación de variables necesarias para el funcionamiento del software.
- Establecer la metodología apropiada para el desarrollo de la herramienta.

- Diseñar la arquitectura para la elaboración de la herramienta.
 - Establecer los requerimientos funcionales y no funcionales de la aplicación.
 - Diseñar el modelado de la herramienta.
 - Seleccionar el algoritmo más adecuado.
 - Inconveniente al momento de obtener las bases de datos por parte de las fuentes necesarias para la investigación.
- Búsqueda y organización de la información de las bases de datos de los estudiantes de Ingeniería de Sistemas.
 - Integración de bases de datos.
 - Realizar pruebas al funcionamiento del software.
 - Analizar el sistema de reglas que permita la extracción del conocimiento.

1.5.2 Delimitación conceptual. Durante la ejecución del proyecto se tendrán en cuenta los siguientes términos como: minería de datos, deserción estudiantil, aprendizaje supervisado, knowledge discovery in databases “descubrimiento de conocimiento en bases de datos” (KDD), extracción del conocimiento, machine learning, java, weka.

1.5.3 Delimitación geográfica. El desarrollo del proyecto tendrá lugar en el programa de ingeniería de sistemas de la universidad Francisco de Paula Santander seccional Ocaña.

1.5.4 Delimitación temporal. En el desarrollo del presente proyecto se emplearán 4 meses a partir de la aprobación de la propuesta de trabajo de grado.

Capítulo 2. Marco referencial

2.1 Marco histórico

2.1.1 Revisión de la deserción estudiantil a nivel internacional. En universidades de países como Chile el fenómeno de la deserción lo han manejado desde dos perspectivas según lo indica un estudio realizado por la universidad metropolitana de ciencias de la educación de Chile, la primera es la interna, que compromete a las IES y se enfoca en los procesos de enseñanza y aprendizaje y la segunda son los procesos externos en relación a lo educativo, en este caso intervienen las características de los estudiantes, el ambiente social en que se encuentran, los métodos de financiación que les brindan a los estudiantes entre otros (Henríquez & Escobar, 2016).

En la universidad de Antofagasta de Chile; Reyes, Escobar, Duarte & Ramírez (2007) implementaron el método de regresión logística, para el cual diseñaron un sistema para poder predecir y conocer cómo les iba a los estudiantes en el primer semestre; para ello se obtuvieron en cuenta factores como el rendimiento, la importancia de acuerdo a su desempeño y la valoración del esfuerzo de cada uno, además de otras variables como: las notas de enseñanza media y los puntajes de las pruebas de selección universitaria (PSU) de lenguaje, matemática y ciencias (Saldaña Villa & Barriga, 2010).

En la universidad Católica de la Santísima Concepción, Chile, utilizaron el modelo de Tinto, Saldaña Villa & Barriga (2010), afirman que: “es un intento por aprovechar al máximo la

información disponible de la universidad para estudiar y comprender la deserción al interior de la carrera de ingeniería civil”. De acuerdo a ello obtuvieron como resultados que tanto el ingreso a la universidad como los ingresos familiares y el puntaje obtenido por el alumno en las Pruebas de Selección Universitaria (PSU) de lenguaje y matemática son las variables que más impactan en la deserción de los estudiantes de la carrera mencionada anteriormente. Posteriormente a ello, luego de que los estudiantes empiezan sus estudios otras variables que afectan a las anteriores son el promedio semestral del alumno, la cantidad de créditos aprobados que el estudiante acumula en cada semestre, y el porcentaje de financiamiento que haya obtenido para cubrir los costos de carrera. De este modo, también se tuvieron en cuenta variables como la residencia de los alumnos y sus preferencias de postulación no resultaron significativas, pero estas, fueron de apoyo al modelo establecido, dado que la dirección de los efectos de éstas sobre la variable dependiente, son acordes con lo hipotetizado previamente (Tinto, Saldaña Villa & Barriga, 2010).

Debido al aumento en el número de estudiantes no solo en la educación presencial sino también en la virtual, los investigadores, instructores y administradores buscan métodos efectivos para extraer conocimiento significativo (Rosado & Verjel, 2017) teniendo en cuenta la minería de datos como procesamiento de los grandes conjuntos de datos (Yukselturk, Ozekes, & Türel, 2014). Actualmente, para esto se encuentran en desarrollo diferentes sistemas que hagan un poco más fácil este proceso. Entre las practicas más importantes para tratar la deserción estudiantil es el identificar de forma temprana cualquier foco del mismo (Mishra, Kumar, & Gupta, 2014), para esto, es necesario detectar a aquellos estudiantes que requieren atención

especial, minimizar el índice de fracaso y tomar medidas aceptables para el próximo examen semestral (Márquez-Vera et al., 2016).

Por otro lado, los índices de deserción en Estados Unidos han ido en aumento, haciendo que la persistencia de estudiantes en carreras de cuatro años sea poco segura, siendo un índice que abarca a los países en general. Estos datos son brindados por medio de las entidades gubernamentales que analizan la deserción. En la India, la tasa de deserción fue del 15% en Bélgica, el índice de deserción fue de 26, el 9% en los Estados Unidos alcanzó el 24%. Según un estudio realizado por la Organización de las Naciones Unidas (ONU) en países latinoamericanos como Colombia y Ecuador, la deserción superó el 40%, mientras que en Costa Rica la deserción alcanzó el 50% y en Brasil inclusive el 54% aproximadamente (Mayra & Mauricio, 2018).

2.1.2 Revisión de la deserción estudiantil a nivel nacional. Un estudio realizado por la universidad de Antioquia, a partir de 1970 se comenzaron a desarrollar las bases conceptuales para estudiar este problema y a finales de la década de los noventa se iniciaron los primeros análisis estadísticos en este campo (Castaño, Gallón, Gómez & Vásquez (2004). Entonces diferentes organizaciones e instituciones han venido trabajando y desarrollado diferentes estrategias para mitigar el número de desertores en las IES, pues su foco fue saber que cada vez aumentaba este número, y tenían que hacer algo al respecto siendo las entidades competentes.

Algo muy importante y que hay que resaltar fue que, gracias al esfuerzo y dinamismo de las instituciones de educación superior, el MEN logro alcanzar un 35.5% de cobertura, tasa que supera el promedio latinoamericano (Guzmán Ruiz, Durán Muriel, & Franco Gallego, 2009) ,

pero con un punto en contra, que los estudiantes estaban desertando, un ejemplo de ello en la universidad de Caldas, poco más de la mitad de los estudiantes (53%) que ingresan a las IES colombianas no culminan estudios y la universidad de Caldas no es ajena a tal situación, 43,6% para abril 2015 (Gartner Isaza & Gallego Giraldo, 2015). Para el año 2009 el MEN publicó el libro “Deserción estudiantil en la educación superior colombiana” en donde se concentra en una metodología de seguimiento, diagnóstico y elementos para su prevención, siendo el apoyo y pilar para las IES ya que suministra información de la situación actual para ese entonces, años anteriores al 2009, y presenta elementos útiles para estudios posteriores frente al análisis y tratamiento de este fenómeno (Guzmán Ruiz, Durán Muriel, & Franco Gallego, 2009).

El MEN también desarrollo un sistema informático: El Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior (SPADIES), el cual consolida y ordena información que permite hacer seguimiento a las condiciones académicas y socioeconómicas de los estudiantes que han ingresado a la educación superior en el país. El objetivo de SPADIES es conocer el estado y evolución de la caracterización y del rendimiento académico de los estudiantes, lo cual es “útil para establecer los factores determinantes de la deserción, para estimar el riesgo de deserción de cada estudiante y para diseñar y mejorar las acciones de apoyo a las estudiantes orientadas a fomentar su permanencia y graduación” (MEN, 2014).

De esta manera las IES, han podido realizar seguimiento a los estudiantes desertores y plantear diferentes variables, ya que los casos no ocurren por el mismo motivo, y así solidificar estrategias que permitan la retención de estudiantes. Por ello es necesario indagar

sistemáticamente e involucrar las diferentes variables, y saber cómo influye una dentro de la otra y así tomar las medidas correctas.

De acuerdo a lo que plantea el ICFES (Universidad de los Andes, 2014) el concepto de deserción puede estudiarse desde tres puntos de vista: (i) individual, (ii) institucional y (iii) estatal o nacional, involucrando a estudiantes, funcionarios de las IES y los responsables de la política nacional de educación. En Colombia, para el año 2016, la deserción en el nivel universitario alcanzó el 46,1 por ciento, lo que significa que uno de cada dos estudiantes que ingresan a la educación superior no finaliza sus estudios (MEN, 2017), con base a este dato las IES tratan de establecer un modelo que permita bajar este valor considerando la interacción entre el sistema académico y social de la universidad (Reyes, Carranza, & Castro, 2012).

Teniendo en cuenta lo anterior, se observa que hay estudios como el realizado por la universidad de la Sabana donde se establecen ciertas estrategias para manejar la deserción teniendo en cuenta algunos patrones identificados por el programa en estudio, en donde se afirma que las tasas de deserción van en aumento en las diferentes instituciones sin importar si son públicas o privadas (Vélez, Amparo; López Jiménez, 2009). Esto, involucra un factor que tienen muy en común las IES públicas y privada y es la dificultad en el aprendizaje que los estudiantes, presente para estudiar y comprender textos, afectando de esta forma en el área académica que puede llegar a la deserción en esos estudiantes (Reyes et al., 2012).

La universidad Industrial de Santander (UIS) aproximadamente en el 2013-2014 implemento programas y estrategias basadas en mejorar la calidad de la educación fomentando el dialogo entre estudiantes y directivos, desde el tema de Responsabilidad Social Universitaria

(RSU) para que las instituciones educativas impacten positivamente a nivel social, económico y ambiental para poder disminuir el nivel de deserción (Zarate & Manilla, 2014).

Otra de las universidades que ha expuesto artículos en donde analizan este tema es la universidad de Antioquia. Un estudio realizado por Castaño, Gallón, Gómez & Vásquez (2004), concluyen que en Colombia son pocas las investigaciones que han intentado estudiar y cuantificar los diferentes factores por los cuales un estudiante decide abandonar sus estudios superiores, e igualmente, formular posibles políticas o reformas educativas que permitan la permanencia de estos, dentro del sistema de educación superior o de la institución universitaria, siendo el común denominador el estudio de las características de la población desertora, la construcción de índices de deserción y la descripción estadística del problema al interior.

Así, posteriormente la misma universidad realizó otro estudio aplicando el método Prentice-Gloeckler y Meyer en dos facultades, Ingeniería y Ciencias Económicas, obteniendo diferentes resultados en donde intervienen características como personales, académicas, socioeconómicas, educación de los padres y adaptación en el ambiente universitario; en cada una de ellas evaluaron y obtuvieron diferentes variables. Por ejemplo, en la parte personal las variables más incidentes son la edad de inicio de los alumnos, personas casadas y personas que provienen de lugares mucho más lejanos, considerando esta última clave para planear estrategias de retención estudiantil. A nivel académico las variables que parecen ser de mayor importancia son el nivel satisfacción con el programa, mayor repetencia de materias, ingreso por segunda opción y no haber hecho cambio de carrera, colegio y experiencia académica. En el aspecto socio-económico depende del estrato y de personas que trabajen, ya que tienen más

responsabilidades a parte de la persona misma y es difícil seguir estudiando por la calidad de vida que carece. Por último, la adaptación del estudiante en la universidad es uno de los factores más influyentes a la hora de desertar (Castaño, Gallón, Gómez, & Vásquez, 2006).

Actualmente la mayoría de los programas académicos de la universidad de Antioquia han realizado estudio sobre la deserción, pero se olvidaron del programa de zootecnia siendo excluidos a las políticas y programas de permanencia, sin embargo, en el 2017 realizaron un estudio el cual concluyeron que se debía prestar más atención lo referente a la afinidad vocacional de los estudiantes, ya que era uno de las causas por las cuales los estudiantes desertaban. Sin embargo, las estrategias que se llevaron a cabo no han sido del todo certeras frente a la mitigación de la problemática real, dificultando avanzar en dicha problemática. Para Gallego, Castrillón, & Marín (2017) es claro que para que un programa de acompañamiento tenga acogida, este debe obedecer a necesidades específicas de la comunidad; por lo tanto, conocer previamente los factores que realmente están asociados con la problemática de la deserción estudiantil en el programa, permitirá definir estrategias de acción para tratar de mitigarlos.

En la universidad de Caldas realizaron un estudio donde analizaban el impacto del apoyo, en factores internos como externos, es decir, culturales, académicos, institucionales y económicos, por parte de las instituciones hacia sus estudiantes. A nivel externo no es que pudieran hacer mucho, pero a nivel interno es la institución quien toma las decisiones y son ellos quienes gobiernan. Por ello la universidad de Caldas ha venido trabajado arduamente para disminuir el porcentaje de deserción vinculando estrategias de apoyo económico, psicológico,

social, familiar, entre otros; articulando lo anterior a un programa llamado “permanencia de calidad”. Finalmente, las variables que inciden a nivel nacional no son diferentes en la universidad de Caldas (Gartner & Gallego, 2015).

2.1.3 Revisión de la deserción estudiantil a nivel regional. A nivel regional las IES han estado trabajando en la búsqueda de una solución que ayude a que los niveles de deserción bajen; estudios que van desde el análisis de la situación fronteriza en Cúcuta como la presencia de grupos armados al margen de la ley que generan escenarios Psicosociales muy difíciles para los estudiantes que están en medio de esta situación, promoviendo la deserción aún más en estos casos (Yolima, Rojas, Pablo, Torres, & Temática, 2016). Además de este tipo de estudios también están los que se toma de una cantidad amplia de datos que permite generar estadísticas y deducir el problema, como es el caso de una tesis realizada en la carrera de administración de la universidad Francisco de Paula Santander Ocaña (UFPSO), dos estudiantes obtuvieron una base de datos que mediante gráficas y estadísticas en Excel lograron obtener una serie de posibles factores que promueven la deserción en el área de administración, tales como falta de motivación y la falta de orientación vocacional estando estas dos opciones entre las más marcadas dentro de la investigación (Gomez & Velasquez, 2017).

Para el estudio estadístico se tienden a tener en cuenta variables como: semestre de retiro, nivel de deserción, nivel académico, estrato social, lugar de residencia, entre otros. Estos, son los datos que se tuvieron en cuenta en un estudio de la Guajira (MELENDEZ, 2008), con el fin de no pasar por alto variables de ningún tipo, tanto externas como internas dentro de la institución.

2.2 Marco contextual

El programa de ingeniería de sistemas de la universidad Francisco de Paula Santander Ocaña presenta un inconveniente en el análisis de la deserción estudiantil, donde en los últimos años la única opción para detectar características dentro de la deserción han sido las consultas una a una dentro de la información que se encuentra en la base de datos del programa. Este problema se ha venido creciendo a lo largo de los años, dado que el incremento de estudiantes en ingeniería de sistemas dificulta cada vez más las consultas manuales que realizan a la base de datos, haciendo necesario el uso de una herramienta que trabaje minería de datos, capaz de detectar esos patrones ocultos que no son percibidos a simple vista.

2.3 Marco conceptual

2.3.1 Deserción estudiantil. Según el ministerio de educación nacional la deserción estudiantil se puede entender como el abandono del sistema escolar por parte de los estudiantes, esto se provoca por la combinación de varios factores que se generan tanto al interior del sistema como en contextos de tipo social, familiar, individual y del entorno. El índice de deserción intra-anual tiene solo en cuenta a los alumnos que abandonan la escuela durante el año escolar (dos semestres consecutivos), ésta se complementa con la tasa de deserción interanual que calcula aquellos que desertan al culminar el año escolar (MINEDUCACION, 2018).

2.3.2 Desertor. Es un alumno que por retiro formal o no formal no está en condiciones de ser evaluado y ha abandonado el sistema educativo (MINEDUCACION, 2018).

2.3.3 Java. Lenguaje de programación orientado a objetos y una plataforma informática comercializada por primera vez en 1995 por Sun Microsystems (JAVA, 2018).

2.3.4 Descubrimiento de conocimiento en bases de datos (KDD). Autores como Fayyad, Piatetsky-Shapiro y Smith (1996) lo definen como “El proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente entendibles al usuario a partir de los datos”. Elementalmente es una metodología en donde interactúan tanto el descubrimiento como el análisis, es decir, metodología en el cual se extrae patrones a partir de datos para posteriormente el usuario los analice.

2.3.5 Minería de datos. La minería de datos se conoce como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información no estructurada (interna y externa a la compañía) en información estructurada, esto, se realiza para su explotación directa o para su análisis y conversión en conocimiento y así dar soporte a la toma de decisiones sobre el negocio (Marcano Aular & Talavera Pereira, 2007), de este modo, la minería de datos ha pasado de ser una herramienta de análisis de información a algo más profundo, como la toma de información de varias fuentes que con el análisis pertinente van a generar los resultados esperado en el futuro (Castorena Peña, Silva Ávila, Domínguez Lugo, & Rodríguez Montelongo, 2018).

2.3.6 Patrones. Según un artículo publicado en la revista scielo los patrones son expresiones que describen una secuencia (Altamiranda, Aguilar, & Hernandez, 2015), esta secuencia se encarga de predecir cuándo va a ocurrir cierto evento con base a las variables que se pretendan estudiar (Timarán Pereira & Yépez Chamorro, 2012).

2.3.7 Permanencia estudiantil. Es conocida como el proceso de integración e interacción individual, social, académica e institucional que admite la mayor duración voluntaria del estudiante dentro de cada año académico en las aulas de clase, para de esta forma, garantizar la terminación de ciclos en los tiempos previstos y asegurando el dominio de las competencias y conocimientos correspondientes (Franzante, 2013).

2.3.8 Sistema de información SPADIES. Es una herramienta por parte del estado que se encarga de hacer seguimiento sobre las cifras de deserción de estudiantes de la educación superior. El seguimiento se da por medio de los datos suministrados por las IES a SPADIES, en este sistema se tienen en cuenta los comportamientos, las causas, variables y riesgos determinantes para desertar. Además, con esta información se agrupan los estudiantes de acuerdo con su riesgo de deserción (SPADIES, 2018).

2.3.9 Software. El software es la parte lógica con la que cuenta un equipo físico (Anaya, 2007), son las instrucciones (programas de cómputo) que cuando se ejecutan proporcionan las características, función y desempeño buscados, es también lo que se conoce como estructuras de datos que permiten que los programas manipulen en forma adecuada la información (Roger S. Pressman, 2010).

2.3.10 Machine learning. El machine learning o aprendizaje automático es definido como un conjunto de métodos que empleados de la forma adecuada con capaces de detectar automáticamente los patrones que se encuentran en los datos (Manament Solutions, 2018).

2.3.11 Weka. Es un conjunto de algoritmos de aprendizaje automático encargado de las tareas de minería de datos. Contiene herramientas para la preparación de datos, clasificación, regresión, agrupación, extracción de reglas de asociación y visualización. En lo que corresponde al software es de código abierto que se encuentra bajo la licencia de GNU/LINUX (WEKA, 2016).

2.4 Marco teórico

2.4.1 Teorías de la deserción estudiantil. Existen diferentes teorías y modelos esenciales frente a la deserción estudiantil, entre ellas:

en 1970 nace la teoría SPADY a cerca de las explicaciones que conllevan a los estudiantes a la deserción; esta teoría es basada en un modelo sociológico de la deserción, el cual explica que el estudiante se encuentra influenciado por antecedentes familiares que afectan sus estados de

integración social en el ámbito universitario. De esta misma forma En 1975 Tinto expande el modelo anteriormente mencionado, Peterson en 1993 expone: “Tinto incorpora la integración social e intelectual de los estudiantes, el cual en donde aquellos estudiantes tengan esta condición tiene mayor probabilidad de continuar en la institución”. Así, si los beneficios de mantenerse en la institución son apreciados por los estudiantes ante el esfuerzo y dedicación, entonces éste permanecerá en la institución. De igual modo, se desprenden diferentes variables para Tinto, entre ellas el nivel socioeconómico, nivel cultural de la familia y los valores de la misma, atributos personales, experiencia postsecundaria y preuniversitaria. Para Tinto los aspectos fundamentales es el sistema social y académico con el que estudiante ingresa a la institución y así determinar de alguna manera el estado de un estudiante (Donoso & Schiefelbein, 2007).

En 1980-1982 aparece Bean el cual incorpora en su modelo el modelo de Tinto, en donde reemplaza las variables del ambiente laboral por las más adecuadas en el entorno de educación superior, es decir, para Bean, la satisfacción con el estudio es similar a la satisfacción con el trabajo. Para este autor, factores como académicos, psicosociales, ambientales y de socialización inciden en esta problemática. A demás resalta que los factores externos a la institución, también afectan significativamente en la deserción (Díaz Peralta, 2008) .

Para 1985 se presenta un modelo multipropósito, Pascarella y Terenzini en 1991 manifiestan que el desarrollo y cambio de los estudiantes es función de cinco conjuntos de variables, los dos primeros están relacionados con antecedentes y características personales a saber: aptitudes, rendimientos, personalidad, aspiraciones, etnicidad. El segundo grupo lo forma las características estructurales y organizacionales de la institución, admisión, estudiantes de la

facultad, selectividad, porcentaje de residentes. Este grupo se relaciona con el tercer grupo: entorno institucional. Dichos grupos afectan al cuarto grupo que considera el conjunto de variables interacciones con los miembros de la facultad y los pares. Por último, el quinto grupo se refiere al esfuerzo del estudiante por aprender constantemente (Donoso & Schiefelbein, 2007, pág. 19). “Estos investigadores concluyen que el ambiente universitario ofrece un escenario en que las oportunidades para el cambio son sustanciales y que quizá sobrepasen a las de cualquier otra institución social” (Gómez Lopez, 2005).

Finalmente, Ethington en 1990 introduce una teoría más general del cual afirma y concluye que “el rendimiento académico previo afecta el desempeño futuro al actuar sobre el modelo conceptual para la deserción estudiantil universitaria chilena auto concepto del estudiante, su percepción de las dificultades de estudio, sus metas, valores y expectativas de éxito” (Ethington, 2008).

La deserción estudiantil es un problema que se viene presentando desde hace años con fuentes diversas dependiendo de la circunstancia y la localización donde se presente. Algunas pueden darse por la desigualdad social en la que se puede definir como la condición de un acceso desigual a ciertos recursos, provocando una fractura en la sociedad entre los que tienen y los que no acceso a esos recursos (Kerbo, Hill, Sistema, Mundial, & Hill, 1991), de este modo, los estudiantes pertenecientes a cierta IES en donde los recursos jueguen un papel importante a la hora de rendir académicamente, ya sea para comprar libros o materiales que sean usados en clase, puede generar un ambiente desmotivador por quienes no pueden obtener estos recursos. Pero el caso no es solo hacer un estudio sobre cierta variable que cause ese efecto, si no de que a

lo largo de la vida universitaria se vayan obteniendo datos que permitan conocer cuando un estudiante puede estar en un riesgo de desertar (Lopera, 2008).

En un estudio realizado por la Universidad de Concepción de Chile se establece que la deserción en el nivel de educación es fatal y no solo corresponde a la decisión que toma el individuo como tal si no también del papel que juega la sociedad que lo rodea para que él tome esta decisión (Quintela-Dávila, 2013), es por eso importante disminuir el índice de los casos de deserción que se presenten dado que al momento de un estudiante de educación superior desertar se está cortando el camino hacia una meta preestablecida y a su vez una inversión intelectual para el desarrollo del estado. Por su parte el estado en Colombia ha venido generando estrategias para disminuir la deserción (Ministerio de Educación Nacional, 2014) que van desde promover una mejor convivencia hasta fomentar el uso de las TIC en los procesos, académicos, estos son opiniones que han resultado de los datos que las universidades registran en el SPADIES, dejando de lado muchos datos que puede analizar la misma IES con base en las variables que se presenten en determinada facultad o carrera.

A nivel institucional, se han planteado diferentes teorías de lo que podría generar la deserción estudiantil esto visto a manera institucional y para ser más específicos en la carrera de administración de empresas, donde se establecen características como falta de motivación y la calamidad (enfermedad o lesión grave) (Henríquez & Escobar, 2016) en un estudiante que posiblemente este propenso a desertar.

2.4.2 Tipos de deserción. En una investigación realizada por la universidad tecnológica de Pereira se plantearon dos tipos de deserción, la académica, encargada del rendimiento del estudiante y la que no es académica, asociada a factores socioeconómicos e individuales (BECERRA & IRAL, 2008). Teniendo en cuenta estos dos tipos en lo que corresponde a lo académico es posible saber que depende de la parte individual directamente y respecto a lo no académico se relaciona con lo rodea a su exterior, por ejemplo, la misma institución o la sociedad en general. En busca de una correcta caracterización e la deserción hay estudios que analizan unos tipos de deserción un poco más minuciosa que cuenta a su vez con subtipos, los tipos de deserción de los que trata son la deserción desde una perspectiva transversal cuyos subtipos son: deserción institucional y cambio de programa, deserción institucional: primer, cuarto y séptimo semestre y la deserción institucional por programa académico y deserción desde una perspectiva longitudinal de los cuales los subtipos son: la función de supervivencia y riesgo y el modelo de episodios múltiples (Cristina et al., 2010). En un estudio de la universidad Tecnológica de Bolívar, basta con distinguir entre tres tipos de deserción, la deserción precoz que se da al momento de un estudiante abandonar un programa antes de iniciar habiendo sido aceptado; la deserción temprana se presenta cuando se abandona el programa durante los primeros cuatro semestres y como tercer tipo esta la deserción tardía que se entiende como abandono desde el quinto semestre en adelante (ANGEL & GRANADOS, 2013), y como antes se ha mencionado este es el tipo de deserción más grave pues ya se ha tenido un recorrido casi completo de lo que sería un profesional.

2.4.3 Teorías la minería de datos en la deserción estudiantil. La minería de datos ha sido una de las principales formas mediante la cual grandes volúmenes de información es procesada y tratada para que se pueda extraer la mayor cantidad de conocimiento de la misma, con el tema de deserción estudiantil la minería de datos es importante pues ayuda encontrar patrones en el comportamiento de los datos, donde no solo es importante un par de variables sino un conjunto como tal, que por medio de algoritmos de machine learnin y herramientas como weka para el procesamiento de los datos o herramientas que se han venido desarrollando con el fin de mejorar esta situación que afecta tanto a el estudiante como a su propio entorno.

La deserción estudiantil ha sido estudiada desde hace ya años con el fin de dar una solución clara al respecto. A la solución de este problema se han unido el estudio con base en los datos para el caso particular de la minería de datos que autores como Fayyad, Piatetsky-Shapiro y Smith (1996) lo definen como “el proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente entendibles al usuario a partir de los datos”. En la universidad de Chile se creó un modelo para la automatización del proceso de determinación de riesgo de deserción (Angulo & Sergio, 2012), con esto se pudo concluir que, una solución al problema de la predicción de la deserción estudiantil se da a través de un sistema de soporte a la toma de decisiones (DSS por sus siglas en inglés Decisión Support System). Años más tarde en un estudio realizado en el Instituto Tecnológico Metropolitano de la ciudad de la ciudad de Medellín se logra crear una metodología para la detección temprana de la deserción de los estudiantes del pregrado sistemas de información en la asignatura lógica y programación (Vásquez, 2015).

“En el año 2012 el ministerio de educación nacional planteó tres causas principales de deserción: rendimiento académico, capacidad económica y orientación vocacional” (Suárez, Luz, & Díaz, 2015), así pues, el MEN afirma y rectifica que las variables siguen siendo las mismas desde años anteriores, resaltando que no todas, pero si las más representativas.

2.5 Marco legal

2.5.1 Capítulo I de la ley general de educación. “Todos los individuos tienen derecho a la educación y, por lo tanto, todos los habitantes del país tienen las mismas oportunidades de acceso al sistema educativo nacional, con sólo satisfacer los requisitos que establecen las disposiciones generales aplicables”.

2.5.2 Artículo 15 de la constitución política de Colombia. En este artículo es establecido el derecho del habeas data, esto en lo que se refiere a ceder modificar o corregir información de los datos que son recolectados por los bancos de datos del estado.

2.5.3 La ley 30 de 1992 en su título V, capítulo III, artículo 117. En este apartado se establece la obligación por parte de las IES y su oficina de bienestar universitario de buscar la forma de encontrar las causas y la solución de la deserción dentro de su plantel, mediante actividades encaminadas al desarrollo físico, psicoafectivo, espiritual y social, es con esta ley que las IES tienen total autonomía para diseñar sus currículos.

2.5.4 Ley general de educación en su artículo 4 establece. “Corresponde al estado, a la sociedad y a la familia velar por la calidad de la educación y promover el acceso al servicio público educativo, y es responsabilidad de la nación y de las entidades territoriales, garantizar su cubrimiento”.

2.5.5 Ley 1188 de 2008. Mediante la cual se regula el registro calificado en las IES en Colombia, esa a su vez hace un seguimiento a las variables asociadas a la deserción, la permanencia y la graduación, así como a las diversas acciones institucionales que apoyan a los estudiantes al respecto.

2.5.6 Ley 1273 de 2009. Por medio de la cual se modifica el código penal, se crea un nuevo bien jurídico tutelado - denominado "de la protección de la información y de los datos"- y se preservan integralmente los sistemas que utilicen las tecnologías de la información y las comunicaciones, entre otras disposiciones.

2.5.7 Artículo 76 de la ley No. 112 de 2011. Dentro de la política de bienestar es necesario identificar las variables que se presentan dentro de la deserción, este es el artículo encargado de hacer cumplir ese mandato, donde se tendrá en cuenta los datos registrados en el ministerio de educación nacional.

2.5.8 La ley estatutaria 1581 de 2012. Esta ley fue creada para la protección de los datos personales en donde prima el derecho constitucional que tienen las personas de saber, actualizar y rectificar la información que es recogida en las bases de datos y archivos, así como de proteger su privacidad.

2.5.9 Ley 1620 de 2013. Mediante esta ley se crea “el sistema nacional de convivencia escolar y formación para el ejercicio de los derechos humanos, la educación para la sexualidad y la prevención y mitigación de la violencia escolar” pieza importante para combatir la deserción estudiantil.

2.5.10 Acuerdo nacional para disminuir la deserción en educación superior. Este es un acuerdo registrado en los años 2013 y 2014 en donde el MEN se compromete a aportar mayor capital en lo relacionado a la deserción y la educación en general, promoviendo los procesos de transferencia en los procedimientos que deben tener las IES para el tratamiento de la deserción dentro de las mismas, este acuerdo se encarga de dar apoyo al ya creado en el 2010 gestionado por el plan nacional de desarrollo nacional para reducir los índices de deserción.

2.5.11 Ley 1712 de 2014. El objeto de la presente ley es regular el derecho de acceso a la información pública, los procedimientos para el ejercicio y garantía del derecho y las excepciones a la publicidad de información.

2.5.12 CONPES 3920 de 2018. El Consejo Nacional de Política Económica Y Social del gobierno nacional (CONPES) el 17 de abril de 2018AAA aprobó el documento CONPES 3920 el que define la política de explotación de datos para el estado colombiano. Este documento deja a Colombia como el primer país en Latinoamérica con una política pública integral que permite el aprovechamiento de los datos para generar desarrollo social y económico.

Este documento deja a las instituciones de educación con una gran puerta hacia la creación, implementación y utilización de herramientas que brinden un aprovechamiento de los datos generados en cada área del conocimiento.

Capítulo 3. Diseño metodológico

3.1 Tipo de investigación

Para la presente investigación se tendrá en cuenta el tipo de investigación descriptiva, ya que este tipo de investigación según Salkind (1999) va a “permitir describir la situación de las cosas en el presente”, con el fin de analizar los datos encontrados y tratados dentro de la investigación con técnicas de minería de datos, en donde se hace uso de la metodología KDD (Knowledge Discovery in Databases) con el fin de encontrar causales que fomentan la deserción dentro de la IES. Este tipo de investigación permite ser la base fundamental para continuar con otros tipos de investigación, ya que es importante describir las características de un grupo para posteriormente analizar esas características desde los diferentes puntos de vista (Salkind, 1999).

3.2 Metodología de la investigación

La metodología que se definió para el desarrollo del proyecto de investigación fue la metodología cualitativa de tipo descriptivo, que contiene las siguientes fases para su completo desarrollo:

- Fase 1: Conocimiento del entorno en el que se pretende desarrollar el proyecto.
- Fase 2: Se encarga del descubrimiento de conocimiento en base de datos (KDD).
- Fase 3: Construcción de la herramienta: en esta fase se desarrolla el software, basándose en el patrón de arquitectura modelo-vista-controlador y al modelo de desarrollo por

capas.

- Fase 5: Evaluación e interpretación: esta fase se encarga de evaluar e interpretar el conocimiento obtenido en la investigación.

3.3 Población

La población objetivo está conformada por los estudiantes de ingeniería de sistemas de la UFPSO.

3.4 Muestra

En la presente investigación, la muestra está conformada por los estudiantes de ingeniería de sistemas de los últimos 14 años (2006-2019).

3.5 Técnicas e instrumentos de recolección

La recolección de información será mediante el uso de base de datos digitales, libros, artículos, estudios previos y videos.

3.6 Análisis de la información

En la presente investigación se desarrolla el análisis de la información bajo el proceso del descubrimiento del conocimiento de bases de datos (KDD), permitiendo interpretar la información para luego establecer el aporte académico: desarrollar un software que permita la

detección de estudiantes desertores del programa de ingeniería de sistemas utilizando procesos y elementos de minería de datos, como un elemento que apoya las estrategias generadas para disminuir los índices de deserción.

Capítulo 4. Resultados

4.1 Conocimiento del entorno en el que se pretende desarrollar el proyecto

En esta fase, se realizaron visitas al plan de estudios de ingeniería de sistemas, con el objetivo de ampliar la perspectiva que se tenía de la deserción estudiantil dentro de la carrera. Gracias a ello, el plan de estudio se comprometió a colaborar con la investigación, logrando a través de ellos admisiones, registro y control, suministraran el set de datos necesario para el desarrollo del proyecto.

4.2 Descubrimiento de conocimiento en bases de datos (KDD)

La presente investigación comenzó de la necesidad de identificar las herramientas, técnicas y atributos apropiados para lograr el objetivo principal, el desarrollo del software que prediga estudiantes posiblemente desertores en el plan de estudio de ingeniería de sistemas de la UFPSO con la ayuda de minería de datos. Para este tipo de necesidad particularmente KDD es un proceso que ayuda a extraer conocimiento de bases de datos, el cual involucra cinco fases: selección de atributos, pre-procesamiento, transformación, minería de datos y evaluación y análisis (Timarán Pereira, Hernández Arteaga, Caicedo Zambrano, Hidalgo Troya, & Alvarado Pérez, 2016); dichas fases permiten integrarse entre ellas para finalmente asistir a los investigadores a la correcta selección de las herramientas, técnicas y atributos.

4.2.1 Selección de datos. En esta fase se estudia y evalúan los datos que son necesarios para la investigación, creando o seleccionando un conjunto de datos sobre el cual se realizara el proceso de descubrimiento.

- Investigar las fuentes de los datos

La fuente de datos para el desarrollo del proyecto es la dependencia del sistema de admisiones, registro y control de la UFPSO, en donde se deben solicitar los datos a través del plan de estudios de ingeniería de sistemas; comprendidos entre el año 2006, año que inicio la carrera, hasta el año 2019 periodo I.

- Establecer las variables a utilizar

Para establecer los atributos que se utilizaron en lo que respecta a la deserción estudiantil, se realizó un artículo llamado: Revisión de técnicas, herramientas, algoritmos y atributos para minería de datos usados en la deserción estudiantil (Ver Apéndice A), dicho artículo contiene más de 50 referencias de artículos y conferencias de diferentes bases de datos, que pretendió encontrar la herramienta, técnica, algoritmo y atributos más utilizados en la deserción estudiantil. En la tabla 1 se muestran los atributos más empleados en la deserción.

Tabla 1
Atributos seleccionados según la literatura

Variables	Descripción
Académicas	
Clase: Estado académico	Atributo que permite conocer el estado del estudiante: activo, inactivo, condicional, reintegrado, entre otros

Código del estudiante	Código que le otorga la universidad de acuerdo a su carrera
Año de admisión a la carrera	Año en que ingreso el estudiante a la Universidad
Promedio acumulado	Promedio, resultado de la suma y división del promedio semestral con la cantidad de semestres
Promedio por semestre	Promedio, resultado de la suma de las notas de la materia por el número de créditos y la división del total de número de créditos
Calificaciones por semestre para cada asignatura	Nota de cada asignatura de acuerdo a las normas establecidas por la institución
Semestre	Semestre en el que va el estudiante de acuerdo a las materias cursadas
Asistencia	Asistencias a las horas de clase
Materias perdidas	Materias que han sido calificadas menores a 3.0
Cantidad de veces que repite cada materia	Número de veces que ha repetido una materia
Materias cursadas	Materias que ha cursado actualmente.
Valoración en las pruebas de estado	En Colombia la entidad es el ICFES el que se encarga de evaluar diferentes conocimientos, y da dicha valoración
Horas dedicadas a estudiar en casa	Horas de estudio dedicadas en casa
Horas de estudio en cada materia	Horas de estudio para cada materia en casa, institución o fuera
Colegio de procedencia en el bachillerato	Colegio donde termino el estudio de bachillerato
Nivel de inglés	Niveles que ha realizado en inglés
Demográficas	
Nombre	Nombre del estudiante
Edad	Edad del estudiante actualmente
Sexo	Sexo: F o M
Estrato	Estrato del estudiante
Ciudad de procedencia	Ciudad de donde proviene el estudiante
Departamento de procedencia	Departamento de donde proviene el estudiante

Dirección	Dirección actual del estudiante
Ocupación	Ocupación laboral del estudiante
Número de hermanos	Número de hermanos
Estado civil	Estado civil del estudiante: casado, soltero, divorciado o viudo
Número de miembros de la familia	Grupo de personas que conviven con el estudiante
Pertenece a una etnia	Etnia indígena a la que pertenece el estudiante
Familiares	
Trabajo del padre	Ocupación del padre
Trabajo de la madre	Ocupación de la madre
Nivel de educación del responsable	Estudios del responsable del estudiante Primaria incompleta, primaria completa, bachillerato incompleto, bachillerato completo, técnico profesional incompleto o completo, tecnólogo profesional incompleto o completo, pregrado, posgrado
Económicas	
Beca académica u otra ayuda	Ayuda estudiantil por parte de la Universidad u otra entidad

Nota. Fuente: Autores del proyecto

- Recopilar los datos

Los atributos mencionados en la tabla 1 fueron suministrados en un archivo formato Excel, llamado: “Datos _sistemas”.

El archivo mencionado anteriormente fue obtenido con 25 atributos y 1395 instancias (ver Apéndice B); los atributos se mencionan en la Tabla 2, correspondientes a información que registran los estudiantes al momento de matricularse y otra se genera durante su permanencia en la universidad como: estado alumno, promedio, entre otras.

Tabla 2
Datos sistemas

Nombre del atributo	Descripción	Cantidad
Estado_alumno	Estado en el que se encuentra el estudiante en el sistema de información.	Activo (399), Inactivo (855), Excluido(141)
Sancion	Sanciones que están detalladas en el acuerdo 065	
Ano_ingreso	Año en que el estudiante ingreso a la universidad	2006(66), 2007 (63), 2008(64), 2009(66), 2010(74), 2011(99), 2012(92), 2013(153),2014(144) , 2015(154), 2016(140),2017(125),2018(95),2019(43)
Código	Código asignado por la universidad a cada estudiante	
Promedio	Promedio acumulado	Rango de 0 a 5
Semestre_actual	Semestre en el que está matriculado el estudiante	
Valoracion_icfes	Con fórmulas matemáticas se mantiene un mismo estándar de calcular la valoración icfes	Rango de 0 a 100
Colegio_procedencia	Es el nombre del colegio de donde el estudiante obtuvo título de bachiller	
Fecha_nacimiento	Es el día, mes y año en el que el estudiante nació	
Sexo	Masculino o femenino	F(400) , M(995)
Estrato_social	Es el nivel de estrato en la que se encuentra el estudiante	1(895), 2(406),3(84),4(9),5(0),6(1)
Lugar_procedencia	Nombre de la ciudad de donde proviene el estudiante	OCANA(606),RIO DE ORO(57),AGUACHICA(81)
Dpto_procedencia	Nombre del departamento de donde proviene el estudiante	NORTE DE SANTANDER(789), CESAR(395)
Direccion	Dirección donde reside el estudiante	
Ocupacion	Ocupación laboral que tiene el estudiante	
Numero_hermanos_estudian	Número de estudiantes que tiene el estudiante	
Estado_civil	SO(Soltero), CA(Casado), UL(Unión libre)	SO(1322),CA(9),SE(5),UL(21) Vacias(38)
Numero_miembros_flia	Número de miembros de la familia	
Desplazado	Estudiante obligado a abandonar su casa o residencia actual	

Poblacion_indigena	Si estudiante pertenece a una etnia indígena
Conflicto	Información para saber si el estudiante ha sido víctima del conflicto armado
Trabajo_padre	Ocupación del padre del estudiante
Trabajo_madre	Ocupación de la madre del estudiante
Nivel_educativo_responsable	Nivel de educación del responsable del estudiante
Beca_academica	Información para conocer si el estudiante tiene algún tipo de beca

Nota. Fuente: **Autores del proyecto**

4.2.2 Pre-procesamiento. En esta fase se pretende garantizar la confiabilidad de la información, es decir, realizar diferentes tareas que determinen la transparencia de los datos.

La información solicitada a la oficina de admisiones, registró y control, se recibió, se analizó y se determinó la clase, dicha clase corresponde a: PROMEDIO, ya que el rendimiento académico es considerado como un tipo de abandono, que contribuye al aumento de la deserción estudiantil en las instituciones de educación superior (Guzmán Ruiz, y otros, 2009). Cabe aclarar que en la literatura la clase era considerada el estado del alumno, pero este atributo no es posible tenerlo en cuenta como clase para la investigación ya que este atributo en la universidad no indica una directa deserción y se cataloga con el mismo valor a diferentes tipos de casos de estudiantes en el caso de los inactivos como es el caso de un estudiante que haya cambiado de pensum o un estudiante que deje de matricular uno o más semestres.

Según Muñoz Reja, Gómez Cejudo, Gualo García, Rivas García, & Piattini Velthuis (Calidad de datos, 2019), exponen los diversos problemas frente a la calidad de los datos en

bases de datos relacionales, algunos de ellos se mencionan en la Tabla 3, siendo este la primera base para descartar datos a nivel de atributos y a nivel de instancias.

A nivel de atributos existe mucha información faltante, dispersa, etc.; por lo que se decidió eliminar 11 atributos, teniendo en cuenta lo que plantean los autores anteriormente mencionados.

Tabla 3
Atributos eliminados

Atributo	Error según autores	# instancias incorrectas
Colegio procedencia	Valores imprecisos	1390
Ocupación	Existencia de sinónimos y valores incorrectos	1337
Dirección	Valores imprecisos, subcadena con contenido invalido para el tipo de dato	54
Numero_miembros_flia	Tuplas semivacías	1395
Desplazado	Tuplas semivacías	1395
Poblacion_indigena	Tuplas semivacías	1350
Conflicto	Tuplas semivacías	1279
Trabajo_padre	Existencia de sinónimos y valores incorrectos	438
Trabajo_madre	Existencia de sinónimos y valores incorrectos	152
Nivel_educativo_responsable	Columna conflictiva	4
Beca_academica	Tuplas semivacías y valores incorrectos	1189

Nota. Fuente: Autores del proyecto

A nivel de instancias se presentaron diferentes errores, cada caso contempla una serie de problemas de calidad de datos, por lo cual para cada uno de ellos se hace la respectiva descripción:

- Violación de valor único: existían nueve códigos duplicados, por lo que son considerados outliers.

- Violación de restricciones de dominio y violación de dominio: para ingresar a la UFPSO, es necesario entregar la documentación, entre ello, el puntaje de la valoración icfes; 70 estudiantes tenían el campo vacío, siendo este obligatorio y 3 estudiantes tenían en el puntaje por fuera del rango establecido.

Atributos como lugar de procedencia y departamento de procedencia se seleccionaron pese a la variación de información, debido a que existían mayor número de estudiantes en ciertas zonas y de esta manera se podrían transformar en la siguiente fase.

Finalmente se hizo necesario crea un nuevo atributo, llamado edad, ya que al tener fecha de nacimiento se obtendría dicha edad y de esta manera dimencionalizar y transformar; ocurrió algo particular en este atributo, se verifico que existía el problema de violación de restricciones, es decir, existían estudiantes menores o igual a 14 años, y estos son consideraron como valores atípicos, por ende, fueron eliminados.

4.2.3 Transformación. Esta fase permite discretizar los atributos para reducir el número de variables y convertir todos los atributos a tipo cadena, dejándolos listos para ser minados. Para ello, existe una fórmula matemática ecuación 1, el cual permite obtener los rangos o intervalos para la discretización de cada atributo, que establece el número de bins en el que se va a discretizar, ya que esto puede afectar en gran medida a la representación y la información brindada por los datos (**Rodríguez Rodríguez, 2010**).

Formula de discretización:

$$Intervalo = \frac{Maximo - Minimo}{\#bins} \quad \text{Ecuación 1}$$

Las tablas desde la 4 a la 14, muestran como fueron discretizados los valores de los diferentes atributos usados.

Tabla 4
Discretización del atributo sanción

SANCIÓN	VALOR
Cualquier tipo de sanción	SI
Campos vacíos	NO

Nota. Fuente: Autores del proyecto

Tabla 5
Discretización del atributo año de ingreso

ANO_INGRESO	VALOR
Entre 2006 y 2010	RG_2006_2010
Entre 2011 y 2015	RG_2011_2015
Entre 2016 y 2020	RG_2016_2020
Entre 2021 y 2025	RG_2021_2025

Entre 2026 y 2030

RG_2026_2030

Años erróneos o nulos

NR

Nota. Fuente: Autores del proyecto

Tabla 6

Discretización del atributo semestre actual

SEMESTRE_ACTUAL	VALOR
1	PRIMERO
2	SEGUNDO
3	TERCERO
4	CUARTO
5	QUINTO
6	SEXTO
7	SEPTIMO
8	OCTAVO
9	NOVENO
10	DECIMO
Datos nulos o vacíos	NR

Nota. Fuente: Autores del proyecto

Tabla 7

Discretización del atributo valoración icfes

VALORACION_ICFES	VALOR
Menor a 30	BAJO
Mayor o igual a 30 y menor a 45	MEDIO_BAJO
Mayor o igual a 45 y menor a 55	MEDIO
Mayor o igual a 55 y menor a 65	MEDIO_ALTO
Mayor o igual a 65 y menor a 70	ALTO
Mayor o igual a 70 y menor o igual a 100	MUY_ALTO

Nota. El rango que se tuvo en cuenta para esta tabla, fue de acuerdo al rango que contempla en Ministerio de Educación Superior. **Fuente:** Autores del proyecto

En el atributo FECHA_NACIMIENTO, se tuvo que realizar varios cambios, primero se agregó una columna “EDAD”, como se lo menciona en la fase 2, dicha columna depende de la fecha de nacimiento; se cambió el mes de inglés a español para que queden en el mismo formato, ya que es en inglés como se los necesitaba. Posteriormente se cambió el orden de la fecha dd-mm-yyyy y se transformó a yyyy-mm-dd. Finalmente, se obtiene la edad y se hace la respectiva discretización.

Tabla 8
Discretización del atributo edad

EDAD	VALOR
Mayor o igual a 14 y menor o igual a 15	RG_14_15
Mayor o igual a 16 y menor o igual a 17	RG_16_17
Mayor o igual a 18 y menor o igual a 19	RG_18_19
Mayor o igual a 20 y menor o igual a 21	RG_20_21
Mayor o igual a 22 y menor o igual a 23	RG_22_23
Mayor o igual a 24 y menor o igual a 25	RG_24_25
Mayor o igual a 26 y menor o igual a 27	RG_26_27
Mayor o igual a 28 y menor o igual a 29	RG_28_29
Mayor o igual a 30 y menor o igual a 31	RG_30_31
Mayor o igual a 32 y menor o igual a 33	RG_32_33
Mayor o igual a 34 y menor o igual a 35	RG_34_35
Mayor o igual a 36 y menor o igual a 37	RG_36_37
Mayor o igual a 38 y menor o igual a 39	RG_38_39
Mayor o igual a 40	RG_40_adelante
Datos nulos o erróneos	NR

Nota. Fuente: Autores del proyecto

Tabla 9
Discretización del atributo estrato social

ESTRATO SOCIAL	VALOR
1	UNO
2	DOS
3	TRES
4	CUATRO
5	CINCO
6	SEIS
Datos vacíos o diferente	NR

Nota. Fuente: Autores del proyecto

Tabla 10
Discretización del atributo lugar de procedencia

LUGAR PROCEDENCIA	VALOR
OCANA	OCANA
Otros lugares	OTRO
Datos vacíos	NR

Nota. Fuente: Autores del proyecto

Para discretizar el atributo DPTO_PROCEDENCIA, se tuvieron en cuenta a los departamentos que tenían mayor número de estudiantes; por ello se decidió trabajar con estudiantes que provienen de los departamentos Cesar y Norte de Santander, 395 y 789 instancias respectivamente.

Tabla 11
Discretización del atributo departamento de procedencia

DPTO_PROCEDENCIA	VALOR
20	CESAR
54	NORTE_DE_SANTANDER
Otros departamentos	OTRO
Datos vacíos	NR

Nota. Fuente: Autores del proyecto

Tabla 12
Discretización del atributo número de hermanos estudiando

NUMERO_HERMANOS_ESTUDIAN	VALOR
0	CERO
1	UNO
2	DOS
3	TRES
4	CUATRO
5	CINCO
6	SEIS
7	SIETE
8	OCHO
9	NUEVE
10	DIEZ
11	ONCE
12	DOCE
13	TRECE
14	CATORCE
15	QUINCE
16	DIESCISEIS
17	DIESCIOCHO
Nulos	NR

Nota. Fuente: Autores del proyecto

Para el atributo ESTADO_CIVIL, ya se consideran cuatro estados: CA (casado), SO (soltero), UL (unión libre), SE (separado); a pesar de ello existían campos vacíos que era necesarios convertirlos a formato nominal como se muestra a continuación:

Tabla 13
Discretización del atributo estado civil

ESTADO_CIVIL	VALOR
vacías	NR

Nota. Fuente: Autores del proyecto

Tabla 14
Discretización del atributo promedio acumulado

PROMEDIO	VALOR
Menor a 0.625	MUY_MALO
Mayor a 0.625 y menor a 1.25	MALO
Mayor a 1.24 y menor a 1.875	MUY_BAJO
Mayor a 1.875 y menor a 2.5	BAJO
Mayor a 2.5 y menor a 3.125	MEDIO_BAJO
Mayor a 3.125 y menor a 3.75	MEDIO
Mayor a 3.75 y menor a 4.375	MEDIO_ALTO
Mayor a 4.375 y menor o igual a 5	ALTO

Nota. Fuente: Autores del proyecto

4.2.4 Minería de datos. En esta fase se tuvieron en cuenta dos secciones: Selección de atributos (ordenamiento y ranqueo) y creación de modelos con los algoritmos propuestos.

4.2.4.1 Selección de atributos con WEKA. El objetivo en este punto fue identificar, mediante el conjunto de datos de entrenamiento ya procesados y transformados, que comprenden 12 atributos y 1262 instancias, aquellos atributos que no aportaban en gran medida a la clase (PROMEDIO), y reducir el número de atributos, de ser necesario, ya que una de las ventajas de trabajar con un conjunto de atributos más pequeño es principalmente que los algoritmos de inducción van a funcionar de una forma más eficiente y los modelos van a tender a ser más comprensibles y probablemente más precisos (García Gutiérrez, 2016).

Gracias a la construcción del artículo (Ver Apéndice A), se realizaron cuatro experimentos mediante la herramienta computacional WEKA, siendo esta el resultado de la investigación de la herramienta más utilizada en minería de datos. Además de validar que los algoritmos J48,

Naive Bayes y RandomizableFileteredClasifier, en lo que respecta a la deserción estudiantil bajo la técnica de clasificación, son los más adecuados para la presente investigación. Dichos experimentos permitirían constatar que los atributos elegidos en la fase 3 serían los más precisos, o por el contrario eliminarlos de no ser así. Para la selección de atributos, WEKA utiliza dos métodos que son necesarios:

- El método de evaluación (Attribute Evaluator): permite determinar la calidad del conjunto de atributos para discriminar la clase: CfsSubsetEval, ClassifierSubsetEval, OneRAttributeEval y CorrelationAttributeEval, fueron los tres evaluadores más eficientes. Además estos evaluadores determinan automáticamente el número de atributos seleccionados y por lo tanto se pueden combinar con la mayoría de métodos de búsqueda.
- El método de búsqueda (Search Method): es la manera de realizar la búsqueda de conjuntos de forma eficiente, existen tres métodos para ello: BestFirts, GreedyStenWise y Ranker.

En la Tabla 15 se muestra el resultado de los cuatro experimentos.

Tabla 15
Selección de atributos con WEKA

Experimento	Algoritmo	Atributo con menor porcentaje	Métodos de búsqueda	Métodos de evaluación
1	No se utiliza	0(0 %) 12 ESTADO_CIVIL	BestFirts	CfsSubsetEval
	J48	0(0 %) 12 ESTADO_CIVIL		
2	Naive Bayes	3(60 %) 12 ESTADO_CIVIL	GreedyStepWise	ClassifierSubsetEval
	Randomizable	2(40 %) 12 ESTADO_CIVIL		
3	No se utiliza	0.02 +- 0.002 8.8 +- 0.75 12 ESTADO_CIVIL	Ranker	OneRAttributeEval
4	No se utiliza	0.0161 12 ESTADO_CIVIL	Ranker	CorrelationAttributeEval

Nota. Fuente: Autores del proyecto

En la tabla anterior se observa al atributo que tuvo menor porcentaje con respecto a los demás (ver Apéndices C, D, E y F). Por lo tanto, se eliminó dicho atributo, ya que era poco concluyente y contradictorio. Finalmente, los atributos que se eligieron se muestran en la siguiente tabla:

Tabla 16

Selección de atributos con métodos y algoritmos de evaluación (ordenamiento y ranqueo)

Atributos finales
ESTADO_ALUMNO
SANCION
ANO_INGRESO
SEMESTRE_ACTUAL
VALORACION_ICFES
EDAD
SEXO
ESTRATO_SOCIAL
LUGAR_PROCEDENCIA
DPTO_PROCEDENCIA
NUMERO_HERMANOS_ESTUDIAN
PROMEDIO (CLASE)

Nota. Fuente: Autores del proyecto

4.2.4.2 Construcción y pruebas del modelo de deserción en la educación superior. Bajo la revisión literaria (Ver Apéndice A), se ha optado por: la inducción de árboles de decisión con el algoritmo J48, ya que representa el conocimiento extraído en un conjunto de reglas de decisión de fácil entendimiento. Así mismo, se eligió la técnica de redes bayesiana al algoritmo Naive Bayes, siendo el segundo más utilizado por su efectividad y eficiencia dentro de la minería de datos. Por último, el tercer lugar de los algoritmos más usados en la predicción está dentro de la técnica meta, Random, en donde en WEKA se divide en tres tipos: RandomizableFilteredClassifier, RandonCommite y RandomSubSpace, siendo el primero el mejor algoritmo, de la técnica meta.

4.2.4.2.1 Algoritmo J48. En este caso, se tiene en cuenta el factor de confianza o nivel de confianza, establecido para la poda del árbol generado, influyendo notoriamente la capacidad de predicción del árbol construido. Para ello se realizaron diferentes pruebas y como resultado se optubo que el valor de confianza mas adecuado es 0.51, ya que a a partir de este valor, el porcentaje de instancias correctamente clasificadas es de 79.3185 %, que por el contrario menores a ese valor, el porcentaje de instancias correctamente clasificadas es de 67.0365 % (ver Tabla 17). Para detallar el modelo generado con el algoritmo J48 ver Apéndice G.

Tabla 17
J48 con factor de confianza

FC de 0.25			FC de 0.51		
=== Summary ===			=== Summary ===		
Correctly Classified Instances %	846	67.0365	Correctly Classified Instances %	1001	79.3185
Incorrectly Classified Instances %	416	32.9635	Incorrectly Classified Instances %	261	20.6815
Kappa statistic	0.5439		Kappa statistic	0.7274	

Mean absolute error	0.1192	Mean absolute error	0.0682
Root mean squared error	0.2441	Root mean squared error	0.1847
Relative absolute error	62.3255 %	Relative absolute error	35.656 %
Root relative squared error	78.9823 %	Root relative squared error	59.7397 %
Total Number of Instances	1262	Total Number of Instances	1262

Nota. Fuente: Autores del proyecto

De esta manera, WEKA también proporciona la siguiente información relacionada con el modelo obtenido llamada matriz de confusión (ver Figura 1), evidenciando el número de instancias clasificadas en cada categoría de la clase promedio.

```

=== Confusion Matrix ===
      a  b  c  d  e  f  g  h  <-- classified as
55   6  1  1  2  3  0  0 |  a = MUY_MALO
 4 144  2  2 33  1  0  0 |  b = BAJO
 2 12 23  1 10  0  0  0 |  c = MALO
 4 19  5 30 12  1  0  0 |  d = MUY_BAJO
12 33  2 10 314 11  2  0 |  e = MEDIO_BAJO
 6  4  1  1 15 384  7  0 |  f = MEDIO
 1  0  0  1  0 29 50  0 |  g = MEDIO_ALTO
 2  1  0  0  0  1  1  1 |  h = ALTO

```

Figura 1. Matriz de confusión J48

A continuación se hace la respectiva análisis de la matriz mencionada anteriormente.

- **MUY MALO:** De 68 instancias, 55 están correctamente clasificadas en el tipo MUY_MALO, 6 instancias se confundieron con el tipo BAJO, 1 instancia se confundió con el tipo MALO y 1 instancia se confundió con el tipo MUY_BAJO, 2 instancias se confundieron con el tipo MEDIO_BAJO y 3 instancias se confundieron con el tipo MEDIO.

- **BAJO:** de 186 instancias, 144 están correctamente clasificadas en el tipo BAJO, 4 instancias se confundieron con el tipo MUY_MALO, 2 instancias se confundieron con el tipo MALO, 2 instancias se confundieron con el tipo MUY_BAJO, 33 instancias se confundieron con el tipo MEDIO_BAJO y 1 instancia se confundió con el tipo MEDIO.

- **MALO:** de 48 instancias, 23 están correctamente clasificadas con el tipo MALO, 2 instancias se confundieron con el tipo MUY_MALO, 12 instancias se confundieron con el tipo MALO, 1 instancia se confundió con el tipo MUY_BAJO y 10 instancias se confundieron con el tipo MEDIO_BAJO.

- **MUY BAJO:** de 71 instancias, 30 están correctamente clasificadas en el tipo MUY_BAJO, 4 instancias se confundieron con el tipo MUY_MALO, 19 instancias se confundieron con el tipo BAJO, 5 instancias se confundieron con el tipo MALO, 12 instancias se confundieron con el tipo MEDIO_BAJO y 1 instancia se confundió con el MEDIO.

- **MEDIO_BAJO:** de 384 instancias, 314 están correctamente clasificadas en el tipo MEDIO_BAJO, 12 instancias se confundieron con el tipo MUY_MALO, 33 instancias se confundieron con el tipo BAJO, 2 instancias se confundieron con el tipo MALO, 10 instancias se confundieron con el tipo MUY_BAJO, 11 instancias se confundieron con el tipo MEDIO Y 2 instancias se confundieron con el tipo MEDIO_ALTO.

- **MEDIO:** de 418 instancias, 384 están clasificadas correctamente en el tipo MEDIO, 6 instancias se confundieron con el tipo MUY_MALO, 4 instancias se confundieron con el tipo

BAJO, 1 instancia se confundió con el MALO, 1 instancia se confundió con el MUY BAJO, 15 instancias se confundieron con el tipo MEDIO_BAJO y 7 instancias se confundieron con el tipo MEDIO_ALTO.

- **MEDIO_ALTO:** de 81 instancias, 50 están clasificadas correctamente en el tipo MEDIO_ALTO, 1 instancia se confundió con el MUY_MALO, 1 instancia se confundió con el tipo MUY_BAJO y 50 instancias se confundieron con el tipo MEDIO_ALTO.

- **ALTO:** de 6 instancias, 1 están clasificadas correctamente ALTO, 2 instancias se confundieron con el tipo MUY_MALO, 1 instancia se confundió con el BAJO, 1 instancia se confundió con el MEDIO, 1 instancia se confundió con el MEDIO_ALTO.

4.2.4.2.2 *Algoritmo Naive Bayes.* Este algoritmo trabaja con la independencia de los atributos entre sí de acuerdo al conjunto de datos y conociendo la clase, en donde el resultado de la clasificación de instancias correctas fue de 56.1807%, contra un 43.8193% de instancias clasificadas incorrectamente (ver Tabla 18). El modelo resultante se observa en el Apéndice G.

Tabla 18
Clasificación de instancias Naive Bayes

Resultados de clasificación de instancias con Naive Bayes		
=== Summary ===		
Correctly Classified Instances	709	56.1807 %
Incorrectly Classified Instances	553	43.8193 %
Kappa statistic	0.4087	
Mean absolute error	0.1335	
Root mean squared error	0.2664	
Relative absolute error	69.8071 %	
Root relative squared error	86.1765 %	

Total Number of Instances 1262

Nota. Fuente: Autores del proyecto

La matriz de confusión para este modelo, se muestra en la figura siguiente:

```

=== Confusion Matrix ===
      a  b  c  d  e  f  g  h  <-- classified as
18 19  0  0 15 15  1  0 | a = MUY_MALO
 2 96  1  4 81  2  0  0 | b = BAJO
 0 24  6  0 18  0  0  0 | c = MALO
 0 40  4  1 22  4  0  0 | d = MUY_BAJO
 3 96  2  0 227 56  0  0 | e = MEDIO_BAJO
27  1  0  2 35 341 12  0 | f = MEDIO
 2  0  0  0  1 58 20  0 | g = MEDIO_ALTO
 1  0  0  0  0  3  2  0 | h = ALTO

```

Figura 2. Matriz de confusión Naive Bayes

- **MUY MALO:** de 68 instancias, 18 están correctamente clasificadas en el tipo MUY_MALO, 19 instancias se confundieron con el tipo BAJO, 15 instancias se confundieron con el tipo MEDIO_BAJO, 15 instancias se confundieron con el tipo MEDIO y 1 instancia se confundió con el tipo MEDIO_ALTO.

- **BAJO:** de 186 instancias, 96 están correctamente clasificadas en el tipo BAJO, 2 instancias se confundieron con el tipo MUY_MALO, 1 instancia se confundió con el tipo MALO, 4 instancias se confundieron con el tipo MUY_BAJO y 81 instancias se confundieron con el tipo MEDIO_BAJO, 2 instancias se confundieron con el tipo MEDIO.

- **MALO:** de 48 instancias, 6 están correctamente clasificadas en el tipo MALO, 24 instancias se confundieron con el tipo BAJO y 18 instancias se confundieron con el tipo MEDIO_BAJO.

- **MUY_BAJO:** de 71 instancias, 1 está correctamente clasificada en el tipo MUY_BAJO, 40 instancias se confundieron con el tipo BAJO, 4 instancias se confundieron con el tipo MALO, 22 instancias se confundieron con el tipo MEDIO_BAJO y 4 instancias se confundieron con el tipo MEDIO.

- **MEDIO_BAJO:** de 384 instancias, 227 están correctamente clasificadas en el tipo MEDIO_BAJO, 3 instancias se confundieron con el tipo MUY_MALO, 96 instancias se confundieron con el tipo BAJO, 2 instancias se confundieron con el tipo MALO y 56 instancias se confundieron con el tipo MEDIO.

- **MEDIO:** de 418 instancias, 341 se clasificaron correctamente en el tipo MEDIO, 27 instancias se confundieron con el tipo MUY_MALO, 1 instancia se confundió con el tipo BAJO, 2 instancias se confundieron con el tipo MUY_BAJO, 35 instancias se confundieron con el tipo MEDIO_BAJO Y 12 instancias se confundieron con el tipo MEDIO_ALTO.

- **MEDIO_ALTO,** de 81 instancias, 20 están clasificadas correctamente en el tipo MEDIO_ALTO, 2 instancias se confundieron con el tipo MUY_MALO, 1 instancia se confundió con el tipo MEDIO_BAJO y 58 instancias se confundieron con el tipo MEDIO.

• **ALTO:** de 6 instancias, 0 instancias están clasificadas correctamente en el tipo ALTO, 1 instancia se confundió con el tipo MUY_MALO, 3 instancias se confundieron con el tipo MEDIO, 2 instancias se confundieron con el tipo MEDIO_ALTO.

4.2.4.2.3 *Algoritmo RandomizableFilteredClassifier*. Este algoritmo depende de un vector de valores aleatorios que son reorganizados en partes iguales, permitiendo ir en línea a la clase dependiente. El número de instancias correctamente clasificadas compete al 92.6307 % siendo uno de los modelos con más porcentaje de acierto (ver Tabla 19). El modelo generado se observa en el Apéndice I.

Tabla 19
Clasificación de instancias con RandomizableFilteredClassifier

Resultados de instancias con RandomizableFilteredClassifier		
=== Summary ===		
Correctly Classified Instances	1169	92.6307 %
Incorrectly Classified Instances	93	7.3693 %
Kappa statistic	0.904	
Mean absolute error	0.0217	
Root mean squared error	0.1015	
Relative absolute error	11.3641 %	
Root relative squared error	32.8476 %	
Total Number of Instances	1262	

Nota. Fuente: Autores del proyecto

La matriz de confusión para este modelo, se muestra en la Figura 3:

```

=== Confusion Matrix ===
      a  b  c  d  e  f  g  h  <-- classified as
62   3  1  0  2  0  0  0 |  a = MUY_MALO
 3 177  2  0  4  0  0  0 |  b = BAJO
 1   8 38  0  1  0  0  0 |  c = MALO
 0   9  4 56  2  0  0  0 |  d = MUY_BAJO
 5  30  2  5 341  1  0  0 |  e = MEDIO_BAJO
 1   0  0  0  4 413  0  0 |  f = MEDIO
 0   0  0  1  0  4  76  0 |  g = MEDIO_ALTO
 0   0  0  0  0  0  0  6 |  h = ALTO

```

Figura 3. Matriz de confusión RandomizableFilteredClassifier

- **MUY_MALO**: de 68 instancias, 62 están correctamente clasificadas en el tipo MUY_MALO, 3 instancias se confundieron con el tipo BAJO, 1 instancia se confundió con el tipo MALO y 2 instancias se confundieron con el tipo MEDIO_BAJO.

- **BAJO**: de 186 instancias, 177 están correctamente clasificadas en el tipo BAJO, 3 instancias se confundieron con el tipo MUY_MALO, 2 instancias se confundieron con el tipo MALO y 4 instancias se confundieron con el tipo MEDIO_BAJO.

- **MALO**: de 48 instancias, 38 están correctamente clasificadas con el tipo MALO, 1 instancia se confundió con el tipo MUY_MALO, 8 instancias se confundieron con el tipo BAJO y 1 instancia se confundió con el tipo MEDIO_BAJO.

- **MUY_BAJO**: de 71 instancias, 56 están correctamente clasificadas en el tipo MUY_BAJO, 9 instancias se confundieron con el tipo BAJO, 4 instancias se confundieron con el tipo MALO y 2 instancias se confundieron con el tipo MEDIO_BAJO.

- **MEDIO_BAJO**: de 384 instancias, 341 están correctamente clasificadas en el tipo MEDIO_BAJO, 5 instancias se confundieron con el tipo MUY_MALO, 30 instancias se confundieron con el tipo BAJO, 2 instancias se confundieron con el tipo MALO, 5 instancias se confundieron con el tipo MUY_BAJO y 1 instancia se confundió con MEDIO.

- **MEDIO**: de 418 instancias, 413 están clasificadas correctamente en el tipo MEDIO, 1 instancia se confundió con el tipo MUY_BAJO y 4 instancias se confundieron con el tipo MEDIO_BAJO.

- **MEDIO_ALTO**: de 81 instancias, 76 están clasificadas correctamente en el tipo MEDIO_ALTO, 1 instancia se confundió con el tipo MUY_BAJO y 4 instancias se confundieron con el tipo MEDIO.

- **ALTO**: de 6 instancias, 6 están clasificadas correctamente en el tipo ALTO.

En síntesis, el mayor porcentaje de instancias correctamente clasificadas se logra con el algoritmo RandomizableFilteredClassifier al igual que el J48. Por ende, los modelos generados son los mejores para su utilización en la fase siguiente. En contraste a ellos, el Naive Bayes presenta un porcentaje por debajo del 60% más sin embargo se lo tendrá en cuenta bajo el resultado de la literatura (Apéndice A).

4.2.4.2.4 *Métodos de evaluación de los modelos.* Para la evaluación de los modelos, se tuvieron en cuenta tres métodos: validación simple, validación cruzada y validación Split; esto con el fin de que los patrones en cada modelo creado sean más precisos. Los resultados se reflejan a continuación.

- Validación simple

Este método, se fundamenta en dividir el conjunto de datos en dos partes, el set de entrenamiento y el set de validación o set de prueba. Los modelos generados con cada uno de los algoritmos, se usan para predecir las nuevas observaciones del set de entrenamiento, obteniendo así la estimación del error de predicción.

El resultado de la validación simple con el conjunto de datos, se evidencia en la tabla 20, el cual permite visualizar las instancias resaltadas como aquellas que fueron las que no coincidieron con el set de entrenamiento.

Tabla 20
Evaluación simple

Modelo J48 con J48 factor en 0,51				Modelo Naïve Bayes				Modelo RandomizableFilteredClassifier			
=== Predictions on test set ===				=== Predictions on test set ===				=== Predictions on test set ===			
inst#	actual prediction	predicted prediction	error	inst#	actual prediction	predicted prediction	error	inst#	actual prediction	predicted prediction	error
1	1:?	1:MUY_MALO	0.75	1	1:?	1:MUY_MALO	0.595	1	1:?	1:MUY_MALO	0.994
2	1:?	1:MUY_MALO	0.75	2	1:?	1:MUY_MALO	0.703	2	1:?	1:MUY_MALO	0.997
3	1:?	2:BAJO	1	3	1:?	5:MEDIO_BAJO	0.575	3	1:?	2:BAJO	0.994
4	1:?	2:BAJO	1	4	1:?	2:BAJO	0.633	4	1:?	2:BAJO	0.994
5	1:?	2:BAJO	0.667	5	1:?	2:BAJO	0.411	5	1:?	3:MALO	0.994
6	1:?	2:BAJO	0.538	6	1:?	2:BAJO	0.611	6	1:?	2:BAJO	0.444
7	1:?	1:MUY_MALO	0.833	7	1:?	1:MUY_MALO	0.93	7	1:?	1:MUY_MALO	0.994
8	1:?	1:MUY_MALO	0.833	8	1:?	1:MUY_MALO	0.894	8	1:?	1:MUY_MALO	0.994

9	1:? 5:MEDIO_BAJO	0.5	9	1:? 2:BAJO	0.594	9	1:? 5:MEDIO_BAJO	0.666
10	1:? 1:MUY_MALO	0.75	10	1:? 1:MUY_MALO	0.703	10	1:? 1:MUY_MALO	0.997
11	1:? 5:MEDIO_BAJO	1	11	1:? 6:MEDIO	0.412	11	1:? 5:MEDIO_BAJO	0.994
12	1:? 1:MUY_MALO	0.833	12	1:? 1:MUY_MALO	0.79	12	1:? 1:MUY_MALO	0.994
13	1:? 6:MEDIO	0.947	13	1:? 6:MEDIO	0.423	13	1:? 6:MEDIO	0.666
14	1:? 6:MEDIO	0.947	14	1:? 6:MEDIO	0.423	14	1:? 6:MEDIO	0.666
15	1:? 6:MEDIO	1	15	1:? 4:MUY_BAJO	0.414	15	1:? 6:MEDIO	0.994
16	1:? 2:BAJO	0.5	16	1:? 2:BAJO	0.562	16	1:? 2:BAJO	0.994
17	1:? 2:BAJO	1	17	1:? 5:MEDIO_BAJO	0.564	17	1:? 2:BAJO	0.994
18	1:? 2:BAJO	0.667	18	1:? 2:BAJO	0.525	18	1:? 2:BAJO	0.994
19	1:? 2:BAJO	1	19	1:? 2:BAJO	0.545	19	1:? 2:BAJO	0.994
20	1:? 2:BAJO	0.5	20	1:? 2:BAJO	0.576	20	1:? 2:BAJO	0.499
21	1:? 5:MEDIO_BAJO	0.75	21	1:? 5:MEDIO_BAJO	0.5	21	1:? 5:MEDIO_BAJO	0.994
22	1:? 2:BAJO	0.7	22	1:? 2:BAJO	0.556	22	1:? 2:BAJO	0.998
23	1:? 2:BAJO	0.667	23	1:? 2:BAJO	0.627	23	1:? 2:BAJO	0.994
24	1:? 2:BAJO	0.538	24	1:? 2:BAJO	0.611	24	1:? 2:BAJO	0.444
25	1:? 2:BAJO	0.333	25	1:? 2:BAJO	0.562	25	1:? 3:MALO	0.4
26	1:? 5:MEDIO_BAJO	0.8	26	1:? 5:MEDIO_BAJO	0.546	26	1:? 5:MEDIO_BAJO	0.994
27	1:? 5:MEDIO_BAJO	0.667	27	1:? 2:BAJO	0.447	27	1:? 2:BAJO	0.499
28	1:? 2:BAJO	0.333	28	1:? 2:BAJO	0.562	28	1:? 3:MALO	0.4
29	1:? 4:MUY_BAJO	0.667	29	1:? 2:BAJO	0.463	29	1:? 3:MALO	0.499
30	1:? 4:MUY_BAJO	0.667	30	1:? 2:BAJO	0.538	30	1:? 4:MUY_BAJO	0.997
31	1:? 5:MEDIO_BAJO	0.625	31	1:? 2:BAJO	0.494	31	1:? 5:MEDIO_BAJO	0.666
32	1:? 2:BAJO	0.333	32	1:? 2:BAJO	0.562	32	1:? 3:MALO	0.4
33	1:? 5:MEDIO_BAJO	0.538	33	1:? 2:BAJO	0.445	33	1:? 5:MEDIO_BAJO	0.994
34	1:? 5:MEDIO_BAJO	0.7	34	1:? 2:BAJO	0.441	34	1:? 2:BAJO	0.5
35	1:? 2:BAJO	0.538	35	1:? 2:BAJO	0.609	35	1:? 2:BAJO	0.749
36	1:? 5:MEDIO_BAJO	0.6	36	1:? 2:BAJO	0.55	36	1:? 3:MALO	0.994
37	1:? 6:MEDIO	1	37	1:? 6:MEDIO	0.455	37	1:? 6:MEDIO	0.994
38	1:? 6:MEDIO	0.947	38	1:? 6:MEDIO	0.423	38	1:? 6:MEDIO	0.666
39	1:? 2:BAJO	0.5	39	1:? 2:BAJO	0.506	39	1:? 5:MEDIO_BAJO	0.994
40	1:? 1:MUY_MALO	0.5	40	1:? 2:BAJO	0.602	40	1:? 1:MUY_MALO	0.499
41	1:? 5:MEDIO_BAJO	1	41	1:? 2:BAJO	0.516	41	1:? 5:MEDIO_BAJO	0.997
42	1:? 1:MUY_MALO	1	42	1:? 2:BAJO	0.553	42	1:? 1:MUY_MALO	0.994
43	1:? 2:BAJO	0.5	43	1:? 5:MEDIO_BAJO	0.559	43	1:? 5:MEDIO_BAJO	0.994
44	1:? 2:BAJO	0.7	44	1:? 2:BAJO	0.461	44	1:? 1:MUY_MALO	0.994
45	1:? 2:BAJO	0.5	45	1:? 2:BAJO	0.551	45	1:? 2:BAJO	0.994
46	1:? 2:BAJO	1	46	1:? 2:BAJO	0.438	46	1:? 2:BAJO	0.994
47	1:? 5:MEDIO_BAJO	0.8	47	1:? 2:BAJO	0.516	47	1:? 5:MEDIO_BAJO	0.994
48	1:? 5:MEDIO_BAJO	0.5	48	1:? 2:BAJO	0.594	48	1:? 5:MEDIO_BAJO	0.666
49	1:? 5:MEDIO_BAJO	1	49	1:? 2:BAJO	0.53	49	1:? 5:MEDIO_BAJO	0.994

Nota. Fuente: Autores del proyecto

Por tanto, el modelo RandomFilteredClassifier es el mejor que se adapta a la técnica de clasificación y el que estima el menor error de predicción. Más sin embargo no se puede dejar atrás al modelo J48, que cuenta con un error de predicción bastante aceptable; finalmente el

modelo Naive Bayes ocupa el último lugar ya que considera un error de predicción en un 50% de los datos.

- Validación cruzada

Este método de evaluación se utiliza cuando la cantidad de datos no es tan grande, dividiendo en n grupos para el conjunto de entrenamiento y $n-1$ para prueba (García Morate). Este evaluador requiere del número de folds, que para el caso se realizó con 5 folds, ya que generar mayor porcentaje en las instancias correctamente clasificadas.

Tabla 21
Evaluación cruzada

CROSS-VALIDATION			
Modelo	Algoritmo	Porcentaje de evaluación	
		Instancias correctamente clasificadas	Instancias incorrectamente clasificadas
Naive Bayes	Naive Bayes	52.6149 %	47.3851 %
Randomizablefilteredclassifier	Randomizablefilteredclassifier	38.2726 %	63.0745 %
J48	J48 (≥ 0.51)	53.7242 %	46.2758 %
	J48 (0.25)	57.9239 %	42.0761 %

Nota. Fuente: Autores del proyecto

Según las pruebas realizadas con WEKA, el mayor porcentaje de instancias correctamente clasificadas fue con el modelo J48 con un factor de confianza de 0.25, pese a eso se observa que la validación cruzada no es uno de los mejores métodos de evaluación.

6	1:?	6:MEDIO	0.917	5	1:?	6:MEDIO	0.616	5	1:?	5:MEDIO_BAJO	0.994
7	1:?	5:MEDIO_BAJO	1	6	1:?	6:MEDIO	0.935	6	1:?	6:MEDIO	0.994
8	1:?	6:MEDIO	0.889	7	1:?	5:MEDIO_BAJO	0.787	7	1:?	5:MEDIO_BAJO	0.994
9	1:?	6:MEDIO	0.769	8	1:?	6:MEDIO	0.724	8	1:?	6:MEDIO	0.994
10	1:?	6:MEDIO	0.917	9	1:?	6:MEDIO	0.616	9	1:?	5:MEDIO_BAJO	0.994
11	1:?	6:MEDIO	0.917	10	1:?	6:MEDIO	0.714	10	1:?	6:MEDIO	0.994
12	1:?	6:MEDIO	0.889	11	1:?	6:MEDIO	0.616	11	1:?	6:MEDIO	0.994
13	1:?	6:MEDIO	0.769	12	1:?	6:MEDIO	0.949	12	1:?	6:MEDIO	0.998
14	1:?	6:MEDIO	0.769	13	1:?	6:MEDIO	0.802	13	1:?	6:MEDIO	0.994
15	1:?	5:MEDIO_BAJO	0.404	14	1:?	6:MEDIO	0.878	14	1:?	6:MEDIO	0.994
16	1:?	5:MEDIO_BAJO	1	15	1:?	5:MEDIO_BAJO	0.75	15	1:?	2:BAJO	0.994
17	1:?	6:MEDIO	0.917	16	1:?	5:MEDIO_BAJO	0.647	16	1:?	5:MEDIO_BAJO	0.994
18	1:?	6:MEDIO	1	17	1:?	6:MEDIO	0.855	17	1:?	6:MEDIO	0.994
19	1:?	5:MEDIO_BAJO	0.667	18	1:?	6:MEDIO	0.73	18	1:?	6:MEDIO	0.994
20	1:?	6:MEDIO	0.769	19	1:?	5:MEDIO_BAJO	0.571	19	1:?	5:MEDIO_BAJO	0.994
21	1:?	7:MEDIO_ALTO	0.667	20	1:?	6:MEDIO	0.616	20	1:?	5:MEDIO_BAJO	0.994
22	1:?	6:MEDIO	0.441	21	1:?	6:MEDIO	0.859	21	1:?	7:MEDIO_ALTO	
23	1:?	6:MEDIO	0.917	22	1:?	6:MEDIO	0.673	22	1:?	6:MEDIO	0.994
24	1:?	6:MEDIO	0.917	23	1:?	6:MEDIO	0.859	23	1:?	6:MEDIO	0.994
25	1:?	6:MEDIO	1	24	1:?	6:MEDIO	0.859	24	1:?	6:MEDIO	0.994
26	1:?	6:MEDIO	0.769	25	1:?	6:MEDIO	0.937	25	1:?	6:MEDIO	0.994
27	1:?	6:MEDIO	0.769	26	1:?	7:MEDIO_ALTO	0.514	26	1:?	6:MEDIO	0.994
28	1:?	5:MEDIO_BAJO	0.6	27	1:?	6:MEDIO	0.878	27	1:?	6:MEDIO	0.994
29	1:?	5:MEDIO_BAJO	0.404	28	1:?	5:MEDIO_BAJO	0.824	28	1:?	6:MEDIO	0.994
30	1:?	5:MEDIO_BAJO	0.615	29	1:?	5:MEDIO_BAJO	0.75	29	1:?	5:MEDIO_BAJO	0.994
31	1:?	2:BAJO	0.5	30	1:?	5:MEDIO_BAJO	0.84	30	1:?	2:BAJO	0.994
32	1:?	6:MEDIO	0.917	31	1:?	5:MEDIO_BAJO	0.874	31	1:?	2:BAJO	0.994
33	1:?	6:MEDIO	1	32	1:?	6:MEDIO	0.954	32	1:?	5:MEDIO_BAJO	0.994
34	1:?	6:MEDIO	0.917	33	1:?	6:MEDIO	0.964	33	1:?	6:MEDIO	0.994
35	1:?	6:MEDIO	1	34	1:?	6:MEDIO	0.845	34	1:?	6:MEDIO	0.994
36	1:?	6:MEDIO	0.917	35	1:?	6:MEDIO	0.936	35	1:?	5:MEDIO_BAJO	0.994
37	1:?	6:MEDIO	0.917	36	1:?	6:MEDIO	0.832	36	1:?	6:MEDIO	0.499
38	1:?	5:MEDIO_BAJO	0.8	37	1:?	6:MEDIO	0.954	37	1:?	6:MEDIO	0.994
39	1:?	6:MEDIO	0.917	38	1:?	5:MEDIO_BAJO	0.859	38	1:?	6:MEDIO	0.994
				39	1:?	6:MEDIO	0.751	39	1:?	5:MEDIO_BAJO	0.994
								39	1:?	6:MEDIO	0.994

Nota. Fuente: Autores del proyecto

De lo anterior, se permite establecer que de los 39 estudiantes que ingresaron en el periodo I de 2019, los señalados en rojo son los estudiantes que tendrían el promedio por debajo de 3.125, el cual permite detectar a cuales estudiantes debería realizar el respectivo seguimiento para establecer que ruta es la adecuada y poder ayudarlos.

4.3 Desarrollo del software

Esta fase se encarga de la creación del software haciendo uso de la metodología Open up,

siendo esta, un proceso unificado mínimamente suficiente, lo que significa que solo se incluye contenido fundamental capaz de adaptarse al contenido o proceso que sea necesario, teniendo como centro la arquitectura para el impulso en el desarrollo (Eclipse, 2007).

4.3.1 Roles.

Tabla 24
Roles

Roles	Encargado
Partes interesadas	Plan de estudios de ingeniería de sistemas de la universidad Francisco de Paula Santander Ocaña
El analista	Brenda Chindoy –Kely Diaz
El arquitecto	Kely Diaz
El desarrollador	Brenda Chindoy –Kely Diaz
El probador	Brenda Chindoy
El Gerente	Alveiro Rosado

Nota. Fuente: Autores del proyecto

4.3.2 Fases. Para el desarrollo de la metodología open up se realizan las siguientes fases:

4.3.2.1 Inicio Fase Iteración. Esta fase se encarga de la comprensión del problema y de esta forma responder a la pregunta: ¿Por qué es necesario el desarrollo de un software que permita la automatización de los procedimientos manuales dentro del proceso KDD? y así, saber que construir e identificar la funcionalidad clave del sistema y de esta forma determinar al menos una posible solución y establecer los riesgos asociados con el proyecto.

4.3.2.2 Fase de Elaboración Iteración. Esta fase se encarga de detallar los requisitos del sistema, la creación de una arquitectura, el diseño de los casos de uso.

4.3.2.2.1 Requisitos funcionales. Usuario:

- El sistema debe permitir registrarse.
- El sistema debe permitir loguearse.
- El sistema debe permitir seleccionar el archivo con los datos de entrenamiento en formato Excel.
- El sistema debe permitir cargar el archivo de entrenamiento en formato Excel a la base de datos.
- El sistema debe permitir la visualización de los atributos de entrenamiento discretizados.
- El sistema debe permitir la representación gráfica en 3D o 2D de cada atributo a través del diagrama de barras.
- El sistema debe permitir exportar los datos de entrenamiento discretizados en formato csv.
- El sistema debe permitir convertir el archivo de entrenamiento de formato csv a arff.
- El sistema debe permitir seleccionar el archivo de entrenamiento en formato arff.
- El sistema debe permitir la creación del modelo.
- El sistema debe permitir la visualización del modelo.
- El sistema debe permitir guardar el modelo.
- El sistema debe permitir cargar el archivo de prueba en formato Excel a la base de datos.
- El sistema debe permitir la visualización de los atributos de prueba discretizados.

- El sistema debe permitir la representación gráfica en 3D o 2D, de cada atributo a través del diagrama de barras.
- El sistema debe permitir exportar los datos de prueba discretizados en formato csv.
- El sistema debe permitir convertir el archivo de prueba csv a arff.
- El sistema debe permitir crear la predicción.
- El sistema debe permitir visualizar la tabla de predicción.
- El sistema debe permitir la representación gráfica en 3D o 2D, de la predicción.
- El sistema debe permitir guardar la predicción.

4.3.2.2.2 Requisitos no funcionales. La aplicación solo está disponible para entornos de escritorio.

- La aplicación solo está disponible en sistemas operativos Windows.
- La aplicación está desarrollada en lenguaje java.
- El usuario tiene un tiempo de aprendizaje máximo de 2 horas.
- El sistema cuenta con manuales estructurados de forma adecuada al usuario.
- El sistema debe emitir mensajes de error o de alerta con el fin de informar y orientar al usuario final.
- La aplicación no podrá ocupar más de 2 GB de espacio en disco.
- El sistema no debe ejecutar ninguna transacción si ocurre un error.

4.3.2.2.3 *Patrón de diseño Modelo-Vista-Controlador (MVC)*. MVC, es un patrón de diseño estandarizado cuyo objetivo es dividir en tres partes la arquitectura del software separando los datos de la interfaz de usuario y colocando como intermediario un controlador que comunica el modelo con la vista (Guzman & Tovar, 2013).

En la figura Patrón de diseño Modelo-Vista-Controlador se pueden observar las divisiones dichas anteriormente, donde la única conexión de la vista con los elementos del modelo es el controlador.

La vista contiene las interfaces que ve el usuario En la parte de controlador se encuentra alojada toda la parte de las acciones que ejerce el usuario la cual conllevan al llamado de acción del modelo o de vista. El modelo representa la parte específica donde la información del sistema opera, en este caso se manejan los paquetes de la fachada, experto, estrategias y DAO (Guzman & Tovar, 2013).

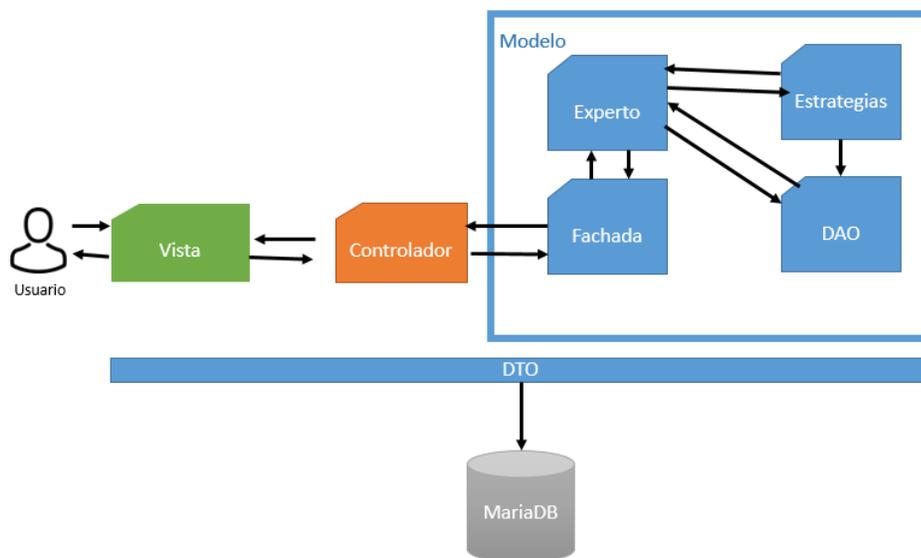


Figura 4. Diagrama Modelo-Vista-Controlador. Autores del proyecto

4.3.2.2.4 Diagrama de actores.

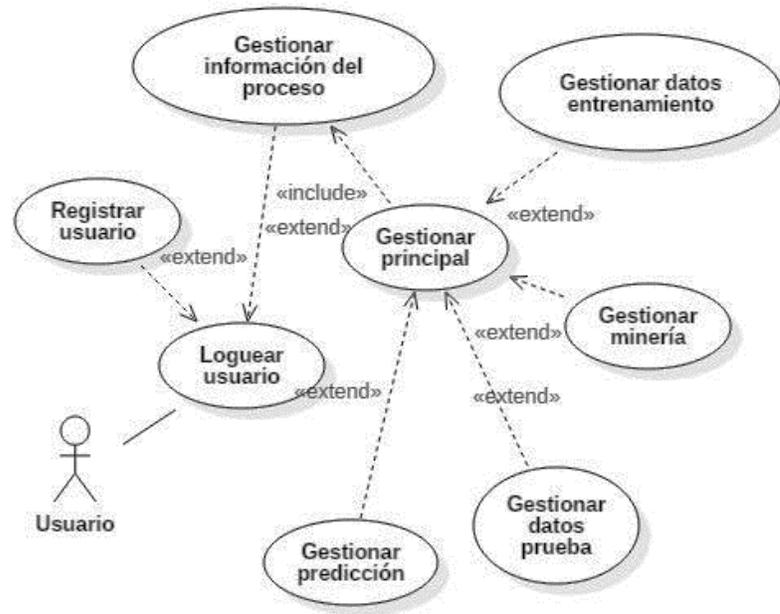


Figura 5. Diagrama de actores. Autores del proyecto

4.3.2.2.4 Diagrama de casos de uso. El diagrama de casos de uso (Tabla 3), describe las acciones del sistema desde el punto de vista del usuario.

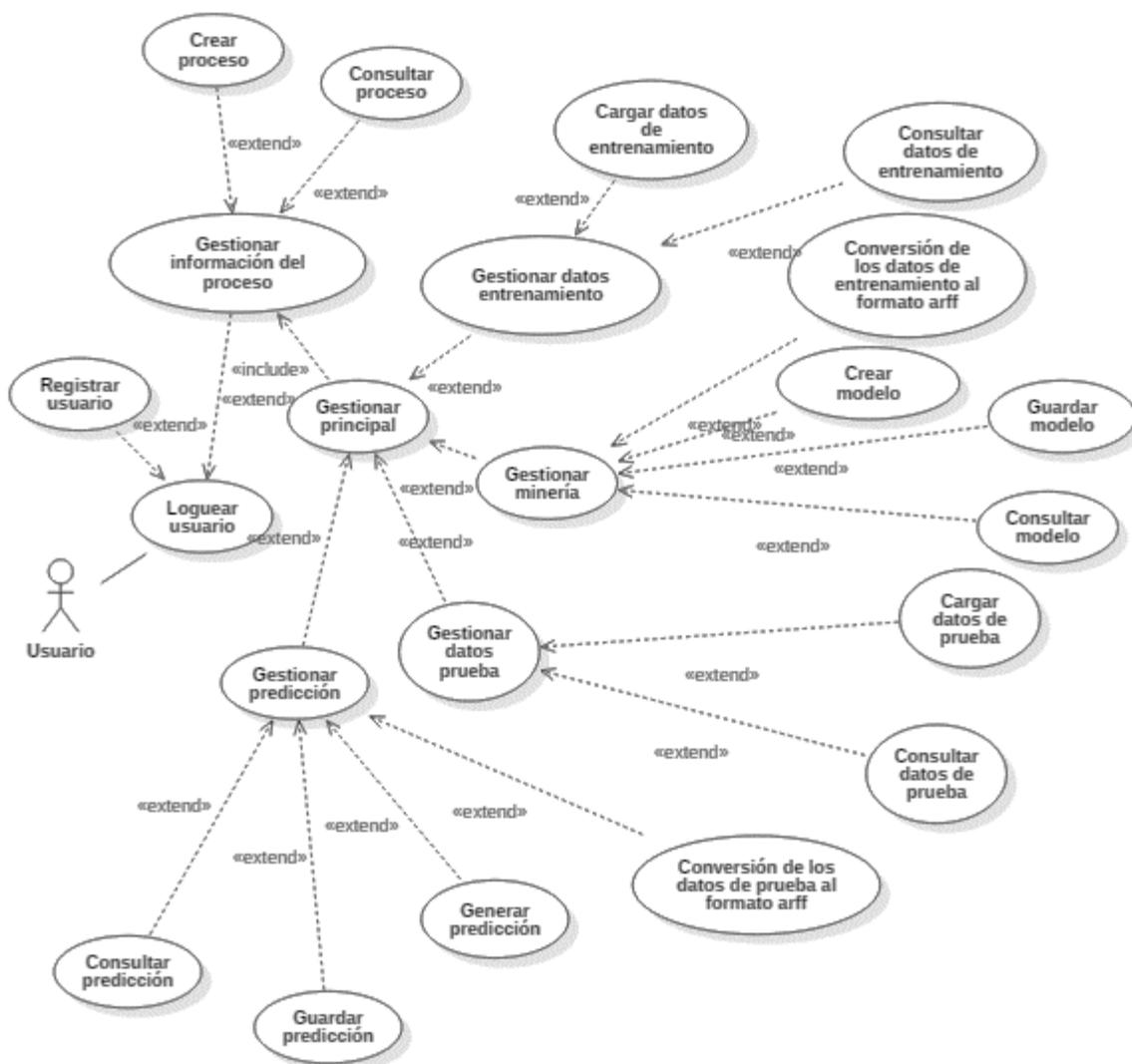


Figura 6. Diagrama de casos de uso. Autores del proyecto

4.3.2.2.5 Especificaciones de casos de uso. Además de la documentación gráfica de las posibles acciones que realiza el actor, se adecua textualmente para una mayor interpretación e identificación de las relaciones existentes con el sistema.

Tabla 25
Registrar usuario

ID - UC:		RU-01
Nombre de UC:		Registrar usuario
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:		Usuario
Descripción:		Permite registrar el usuario al sistema
Disparador:		
Pre-condiciones:		Usuario registrado
Post-condiciones:		
	1	El caso de uso inicia cuando el actor selecciona la opción registrarse
	2	El sistema muestra una ventana con un campo para ingresar la clave maestra
	3	El actor ingresa la clave maestra
	4	El actor selecciona la opción aceptar
	5	El sistema valida campo
Flujo Normal:	6	El sistema muestra un formulario con: código institucional, identificación, nombre, apellido, cargo, teléfono, correo, clave y confirmación de la clave
	7	El actor diligencia el formulario
	8	El actor selecciona la opción registrar
	9	El sistema valida los campos del formulario
	10	El sistema registra el usuario
Flujo Alternativo:	3.1	Si el campo no se ha diligenciado, el sistema muestra un mensaje: <i>"Ingrese clave maestra"</i> . Y vuelve al punto 2.
	3.2	Si la clave maestra no es correcta, el sistema muestra un mensaje: <i>"La clave maestra es incorrecta"</i> . Y vuelve al punto 2.
Excepciones:	9.1	Si los campos no se han diligenciado, el sistema muestra un mensaje: <i>"Hay campos vacíos, debe llenar todos los campos"</i> , y vuelve al punto 5.
	9.2	Si las claves no coinciden, el sistema muestra un mensaje: <i>"Las claves no coinciden"</i> . Y vuelve al punto 5.
	9.3	Si el usuario ya existe, el sistema muestra un mensaje: <i>"El usuario ya existe"</i> . Y vuelve al punto 5.

9.4 Si el email no es válido, el sistema debe mostrar un mensaje: “*El email es incorrecto*”. Y vuelve al punto 5.

Inclusiones:	
Prioridad:	Deseada
Frecuencia de uso:	A veces

Nota. Fuente: Autores del proyecto

Tabla 26
Loguear usuario

ID - UC:		LU-01
Nombre de UC:		Loguear usuario
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:		Usuario
Descripción:		Permite loguearse para ingresar al sistema
Disparador:		
Pre-condiciones:		El usuario debe estar registrado
Post-condiciones:		Usuario logueado
	1	El sistema muestra una interfaz para ingresar el código institucional y clave y en la parte inferior de la pantalla tres botones: registrar, aceptar y cancelar
	2	El actor ingresa en código institucional y la clave
	3	El actor selecciona la opción aceptar
	4	El sistema valida los campos
	5	El sistema permite el ingreso del usuario a la vista de mis proyectos
Flujo Alternativo:	2.1	El actor selecciona la opción registrar y accede al caso de uso RU-01
	2.2	El actor selecciona la opción cancelar y se cierra la ventana
Excepciones:	4.1	Si los campos no se han diligenciado, el sistema muestra un mensaje: “ <i>Debe ingresar sus datos</i> ” y vuelve al punto 2
	4.2	Si los datos son incorrectos, muestra un mensaje: “ <i>Datos incorrectos, vuelva a intentarlo</i> ”. Y vuelve al punto 2.
Inclusiones:		GIU-01
Prioridad:		Esencial
Frecuencia de uso:		Siempre

Nota. Fuente: Autores del proyecto

Tabla 27
Gestionar principal

ID - UC:	GPRI-01	
Nombre de UC:	Gestionar principal	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite realizar todos los procesos del sistema	
Disparador:		
Pre-condiciones:	Debe estar logueado	
Post-condiciones:	Vista principal gestionada	
Flujo Normal:	1	El caso de uso inicia luego de crear o seleccionar un proceso
	2	El sistema muestra la ventana de la vista principal del proyecto con las opciones de archivo, datos entrenamiento, minería, prueba predicción y ayuda
	3	El actor selecciona una opción de la ventana
	4	El sistema permite el acceso a la opción seleccionada
	3.1	El actor selecciona la opción cargar datos de entrenamiento que se encuentra dentro de datos entrenamiento El sistema permite acceso al caso de uso GDE-01
	3.2	El actor selecciona la opción ver datos de entrenamiento que se encuentra dentro de datos entrenamiento y accede al caso de uso GDE-02
	3.3	El actor selecciona la opción convertir datos de entrenamiento a arff que se encuentra en minería y accede al caso de uso GM-01
	3.4	El actor selecciona la opción crear modelo que se encuentra en minería y accede al caso de uso GM-02
	3.5	El actor selecciona la opción ver modelo que se encuentra en minería y accede al caso de uso GM-03
	3.6	El actor selecciona la opción de cargar datos de prueba que se encuentra en prueba y accede al caso de uso GDP-01
Flujo Alternativo:	3.7	El actor selecciona la opción de ver datos de prueba que se encuentra en prueba y accede al caso de uso GDP-02
	3.8	El actor selecciona la opción convertir datos de prueba a arff que se encuentra en minería y accede al caso de uso GDPRE-01
	3.9	El actor selecciona la opción predicción que se encuentra en predicción y accede al caso de uso GDPRE-02
	3.10	El actor selecciona la opción ver predicción que se encuentra en predicción y accede al caso de uso GDPRE-03

3.11 El actor selecciona la opción ayuda y accede al manual del usuario

Excepciones:	
Inclusiones:	Debe crear o abrir un proceso creado
Prioridad:	Esencial
Frecuencia de uso:	Siempre

Nota. Fuente: Autores del proyecto

Tabla 28
Gestionar proceso

ID - UC:	GPR-01	
Nombre de UC:	Gestionar proceso	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite realizar la gestión del proceso	
Disparador:		
Pre-condiciones:	Debe estar logueado	
Post-condiciones:	Vista proceso gestionada	
Flujo Normal:	1	El caso de uso inicia cuando el actor se loguea en el sistema
	2	El sistema muestra la ventana de la vista proceso con una tabla en la que e especifican los proyectos creados y dos botones en la parte inferior que son para seleccionar el proyecto creado o para crear un nuevo proyecto
	3	El actor selecciona una opción de la ventana
	4	El sistema permire el acceso a la opción seleccionada
Flujo Alternativo:	3.1 El actor selecciona un proceso creado y accede al caso de uso GPRI-01	
	3.2 El actor selecciona la opción crear nuevo proceso y accede a el caso de uso GP-01	
Excepciones:		
Inclusiones:		
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

Tabla 29
Crear proceso

ID - UC:	GP-01	
Nombre de UC:	Gestionar información del proyecto (Crear proceso)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite crear un proceso en el sistema	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema	
Post-condiciones:	proceso creado	
	1	El caso de uso inicia cuando el actor selecciona la opción crear proyecto
	2	El sistema muestra un formulario con: fecha de creación, año de inicio, año final y descripción
Flujo Normal:	3	El actor diligencia los campos
	4	El actor selecciona crear proyecto
	5	El sistema crea el proyecto
Flujo Alternativo:		
Excepciones:	4.1 Si los campos no se han diligenciado, el sistema muestra un mensaje: “ <i>Hay campos vacíos, debe llenar todos los campos</i> ”, y vuelve al punto 2	
Inclusiones:	LU-01	
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

Tabla 30
Consultar proceso

ID - UC:	GP-02	
Nombre de UC:	Gestionar información del proyecto (Consultar proceso)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite consultar un proyecto en el sistema	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema	
	Debe tener un proyecto registrado	
Post-condiciones:	Proyecto consultado	

	1	El caso de uso inicia cuando el actor se loguea en el sistema
	2	El sistema muestra una ventana con los proyectos creados
Flujo Normal:	3	El actor elige el proyecto
	4	El actor selecciona la opción ver proyecto
	5	El sistema valida que seleccione proyecto
	6	El sistema muestra ventana principal con el proyecto seleccionado
Flujo Alternativo:		
Excepciones:	5.1	Si no se ha seleccionado ningún proceso, el sistema muestra el mensaje: “No ha seleccionado ningún proyecto”, y vuelve al punto 2
Inclusiones:		LU-01 GP-01
Prioridad:		Deseada
Frecuencia de uso:		A veces

Nota. Fuente: Autores del proyecto

Tabla 31
Cargar datos entrenamiento

ID - UC:	GDE-01	
Nombre de UC:	Gestionar datos entrenamiento (Cargar datos de entrenamiento)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite cargar los datos de entrenamiento	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Debe estar creado un proyecto	
Post-condiciones:	Datos de entrenamiento cargados El caso de uso inicia cuando el actor selecciona la opción cargar datos de entrenamiento	
Flujo Normal:	1	
	2	El sistema muestra la interfaz del cargue de datos de entrenamiento con la opción: buscar archivo de entrenamiento

3	El actor selecciona el botón buscar archivo de entrenamiento
4	El sistema muestra una interfaz para buscar el archivo de entrenamiento
5	El actor selecciona el archivo de entrenamiento
6	El actor selecciona la opción cargar archivo de entrenamiento
7	El sistema comprueba la validez del campo
8	El sistema carga los datos de entrenamiento en la base de datos

Flujo Alternativo:

Excepciones: 7.1 Si no ha seleccionado ningún archivo, el sistema debe mostrar un mensaje: “*Seleccione archivo de entrenamiento*” y vuelve al punto 2.

Inclusiones: LU-01
GIP-01

Prioridad: Deseada

Frecuencia de uso: A veces

Nota. Fuente: Autores del proyecto

Tabla 32

Consultar la información de los datos de entrenamiento

ID - UC:	GDE-02	
Nombre de UC:	Gestionar datos entrenamiento (Consultar datos de entrenamiento)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite consultar los datos de entrenamiento	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Los datos de entrenamiento deben estar cargados	
Post-condiciones:	Datos de entrenamiento consultados	
	1	El caso de uso inicia cuando el actor selecciona la opción ver datos de entrenamiento
Flujo Normal:	2	El sistema muestra una interfaz que permite ver los datos de entrenamiento la cual tiene una lista desplegable que muestra los atributos y la opción de graficar
	3	El actor selecciona el atributo en la lista

4	El sistema realiza la búsqueda
5	El sistema muestra la información del atributo seleccionado en la lista
6	El actor selecciona la dimensión en la que desea graficar: 2D o 3D
7	El actor selecciona la opción graficar
8	El sistema muestra la gráfica en la dimensión que se seleccionó
Flujo Alternativo:	
Excepciones:	7.1 Si no se selecciona una dimensión el sistema muestra un mensaje: “ <i>Seleccione una dimensión</i> ” y vuelve al paso 2
Inclusiones:	LU-01 GDE-01
Prioridad:	Deseada
Frecuencia de uso:	A veces

Nota. Fuente: Autores del proyecto

Tabla 33
Convertir los datos entrenamiento al formato arff

ID - UC:	GM-01	
Nombre de UC:	Gestionar minería (Conversión de los datos de entrenamiento al formato arff)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite convertir los datos de entrenamiento al formato arff	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Los datos de entrenamiento deben estar cargados	
Post-condiciones:	Datos de entrenamiento convertidos al formato arff	
Flujo Normal:	1	El caso de uso inicia cuando el actor selecciona la opción convertir los datos de entrenamiento a arff
	2	El sistema muestra una interfaz de la conversión de los datos de entrenamiento al formato arff, que tiene las opción de : exportar datos de entrenamiento a formato csv, seleccionar el archivo con datos de entrenamiento en formato csv
	3	El actor selecciona la opción de exportar datos de entrenamiento a formato csv

4	El sistema muestra una tabla con los datos de entrenamiento exportados en csv
5	El actor selecciona la opción buscar datos de entrenamiento en formato csv
6	El sistema muestra una interfaz para buscar la ruta del archivo de entrenamiento en formato csv
7	El actor selecciona el archivo en formato csv
8	El sistema obtiene el archivo
9	Validar archivo csv
10	El actor selecciona la opción convertir a formato arff
11	El sistema hace la conversión al formato arff
12	El sistema guarda el archivo de entrenamiento en formato arff

Flujo Alternativo:

Excepciones: 12.1 Si no se selecciona una ruta, el sistema debe mostrar un mensaje: “*Debe especificar el archivo csv*” y vuelve al punto 4

Inclusiones: LU-01
GDE-02

Prioridad: Deseada

Frecuencia de uso: A veces

Nota. Fuente: Autores del proyecto

Tabla 34
Crear modelo

ID – UC:	GM-02	
Nombre de UC:	Gestionar modelo (Crear modelo)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite crear el modelo	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Debe estar creado un proyecto Debe estar el archivo de entrenamiento en formato arff	
Post-condiciones:	Modelo creado	
Flujo Normal:	1	El caso de uso inicia cuando el actor selecciona la opción crear modelo

	2	El sistema muestra una interfaz para la creación del modelo con las opción es: ruta del archivo de entrenamiento y algoritmo de minería de datos
	3	El actor selecciona el archivo de entrenamiento
	4	El actor selecciona el algoritmo
	5	Validación de información
	6	El sistema crea el modelo al seleccionar el algoritmo
	7	El sistema muestra el modelo generado
Flujo Alternativo:		
Excepciones:		6.1 Si no ha seleccionado la ruta con el archivo de entrenamiento, el sistema debe mostrar un mensaje: <i>“Debe seleccionar los datos de entrenamiento para generar el modelo”</i> y vuelve al punto 2.
Inclusiones:		LU-01 GIP-01 GDE-04
Prioridad:		Deseada
Frecuencia de uso:		A veces

Nota. Fuente: Autores del proyecto

Tabla 35
Guardar modelo

ID – UC:	GM-03	
Nombre de UC:	Gestionar modelo (Guardar modelo)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite guardar el modelo	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Crear modelo	
Post-condiciones:	Modelo guardado	
	1	El caso de uso inicia cuando el actor selecciona la opción guardar modelo
Flujo Normal:	2	El sistema muestra una interfaz para guardar la modelo
	3	El actor diligencia los campos con los datos requeridos

	4	El actor selecciona la opción guardar modelo
	5	Validar información
	6	El sistema guarda el modelo
Flujo Alternativo:		
Excepciones:	6.1	Si los datos no están completos, el sistema debe mostrar un mensaje: “ <i>Diligencia todos los campos obligatorios</i> ” y vuelve al punto 2.
Inclusiones:		LU-01 GM-01
Prioridad:		Deseada
Frecuencia de uso:		A veces

Nota. Fuente: Autores del proyecto

Tabla 36
Consultar modelo

ID - UC:	GM-04	
Nombre de UC:	Gestionar modelo (Consultar modelo)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	22 de octubre 2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite generar el modelo	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Guardar modelo	
Post-condiciones:	Modelo consultada	
	1	El caso de uso inicia cuando el actor selecciona la opción consultar modelo
	2	El sistema muestra una interfaz para consultar el modelo
Flujo Normal:	3	El actor selecciona el modelo
	4	El sistema muestra la modelo en un archivo excel
Excepciones:		
Inclusiones:		LU-01 GM-02
Prioridad:		Deseada
Frecuencia de uso:		A veces

Nota. Fuente: Autores del proyecto

Tabla 37
Cargar datos de prueba

ID - UC:	GDP-01	
Nombre de UC:	Gestionar datos prueba (Cargar datos de prueba)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite cargar los datos de prueba	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Debe estar creado el proyecto	
Post-condiciones:	Datos de prueba cargados	
Flujo Normal:	1	El caso de uso inicia cuando el actor selecciona la opción cargar datos de prueba
	2	El sistema muestra la interfaz del cargue de datos de prueba con la opción: buscar archivo de prueba
	3	El actor selecciona la opción buscar archivo de prueba
	4	El sistema le muestra una interfaz para buscar la ruta del archivo de prueba
	5	El actor selecciona el archivo de prueba
	6	El actor selecciona la opción cargar archivo de prueba
	7	El sistema comprueba la validez del campo
	8	El sistema carga los datos de prueba en la base de datos
Flujo Alternativo:		
Excepciones:	7.1 Si no ha seleccionado ningún archivo, el sistema debe mostrar un mensaje: “ <i>Seleccione archivo de prueba</i> ” y vuelve al punto 2.	
Inclusiones:	LU-01 GIP-01	
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

Tabla 38
Consultar la información de los datos de prueba

ID - UC:	GDP-02	
Nombre de UC:	Gestionar datos prueba (Consultar datos de prueba)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite consultar los datos de prueba	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Los datos de prueba deben estar cargados	
Post-condiciones:	Datos de prueba consultados	
Flujo Normal:	1	El caso de uso inicia cuando el actor selecciona la opción ver datos de prueba
	2	El sistema muestra una interfaz que permite ver los datos de prueba la cual tiene una lista desplegable que muestra los atributos: estado del alumno, sanción, año de ingreso, promedio, semestre actual, valoración ICFES, fecha de nacimiento, sexo, estrato social, lugar de procedencia, departamento de procedencia y número de hermanos estudiando y la opción de graficar
	3	El actor selecciona el atributo en la lista
	4	El sistema muestra la información del atributo seleccionado en la lista
	5	El actor selecciona la dimensión en la que desea graficar
	6	El actor selecciona la opción graficar
	7	El sistema muestra la gráfica en la dimensión que se seleccionó
Flujo Alternativo:		
Excepciones:	7.1 Si no se selecciona una dimensión el sistema muestra un mensaje: “ <i>Seleccione una dimensión</i> ” y vuelve al paso 2	
Inclusiones:	LU-01 GDP-01	
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

Tabla 39
 Convertir los datos de prueba al formato arff

ID - UC:	GDPRE-01	
Nombre de UC:	Gestionar predicción (Conversión de los datos de prueba al formato arff)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite convertir los datos de prueba al formato arff	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Los datos de prueba deben estar cargados	
Post-condiciones:	Datos de prueba convertidos al formato arff	
Flujo Normal:	1	El caso de uso inicia cuando el actor selecciona la opción convertir los datos de prueba al formato arff
	2	El sistema muestra una interfaz de la conversión de los datos de prueba al formato arff, que tiene las opción de : exportar datos de prueba a formato csv, seleccionar el archivo con datos de prueba en formato csv
	3	El actor selecciona la opción de exportar datos de prueba a formato csv
	4	El sistema le muestra una tabla con los datos de prueba exportados en csv
	5	El actor selecciona la opción buscar datos de prueba en formato csv
	6	El sistema le muestra una interfaz para buscar la ruta del archivo de prueba en formato csv
	7	El actor selecciona el archivo
	8	El actor selecciona la opción convertir a formato arff
	9	El sistema hace la conversión al formato arff
	10	Validar información
	11	El sistema guarda el archivo en formato arff
Flujo Alternativo:		
Excepciones:	11.1 Si ocurre un error en la exportación de los datos, el sistema debe mostrar un mensaje: “ <i>Debe especificar la ruta del archivo csv</i> ” y vuelve al punto 2.	
Inclusiones:	LU-01 GDP-02	
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

Tabla 40
Crear predicción

ID - UC:	GUPRE-02	
Nombre de UC:	Gestionar predicción (Crear predicción)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite crear la predicción	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema Crear modelo	
Post-condiciones:	Conversión de los datos de prueba al formato arff Predicción creada	
Flujo Normal:	1	El caso de uso inicia cuando el actor selecciona la opción crear predicción
	2	El sistema muestra una interfaz para crear la predicción que contiene datos como: ruta del archivo de prueba y ruta del archivo del modelo
	3	El actor diligencia los campos con los datos requeridos
	4	El actor selecciona la opción generar predicción
	5	Validar información
	6	El sistema genera la predicción
	7	El actor selecciona la dimensión en la que desea graficar
	8	Validar dimensión
	9	El sistema muestra la gráfica en la dimensión que se seleccionó
Flujo Alternativo:		
Excepciones:	6.1 Si los datos no están completos, el sistema debe mostrar un mensaje: <i>“Diligencia todos los campos obligatorios”</i> y vuelve al punto 2.	
	9.1 Si no se selecciona una dimensión el sistema muestra un mensaje: <i>“Seleccione una dimensión”</i> y vuelve al paso 2	
Inclusiones:	LU-01 CM-01	

Prioridad:	GDP-04
Frecuencia de uso:	Deseada
	A veces

Nota. Fuente: Autores del proyecto

Tabla 41
Guardar predicción

ID - UC:	GUPRE-03	
Nombre de UC:	Gestionar predicción (Guardar predicción)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite guardar el predicción	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema	
Post-condiciones:	Crear predicción	
	Predicción guardada	
	1	El caso de uso inicia cuando el actor selecciona la opción guardar predicción
	2	El sistema muestra una interfaz para guardar la predicción
	3	El actor diligencia los campos con los datos requeridos
Flujo Normal:	4	El actor selecciona la opción guardar predicción
	5	Validar información
	6	El sistema guarda la predicción
Flujo Alternativo:		
Excepciones:	6.1 Si los datos no están completos, el sistema debe mostrar un mensaje: <i>"Diligencia todos los campos obligatorios"</i> y vuelve al punto 2.	
Inclusiones:	LU-01	
	GUPRE-02	
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

Tabla 42
Consultar predicción

ID - UC:	GUPRE-04	
Nombre de UC:	Gestionar predicción (Consultar predicción)	
Creado por:	Brenda-Kely	Actualizado por:
Fecha de la creación:	11/06/2019	Fecha de la última actualización:
Actores:	Usuario	
Descripción:	Permite generar el predicción	
Disparador:		
Pre-condiciones:	El usuario debe estar logueado en el sistema	
Post-condiciones:	Guardar predicción	
		Predicción consultada
		El caso de uso inicia cuando el actor selecciona la opción consultar predicción
Flujo Normal:	1	El sistema muestra una interfaz para consultar la predicción
	2	El actor selecciona la predicción
	3	El sistema muestra la predicción en un archivo excel
	4	
Excepciones:		
Inclusiones:	LU-01 GUPRE-03	
Prioridad:	Deseada	
Frecuencia de uso:	A veces	

Nota. Fuente: Autores del proyecto

4.3.2.2.6 Diagrama de actividades.

En los diagramas de actividades se ilustra la dinámica del sistema y el usuario mediante el modelado del flujo ocurrente de actividad en actividad.

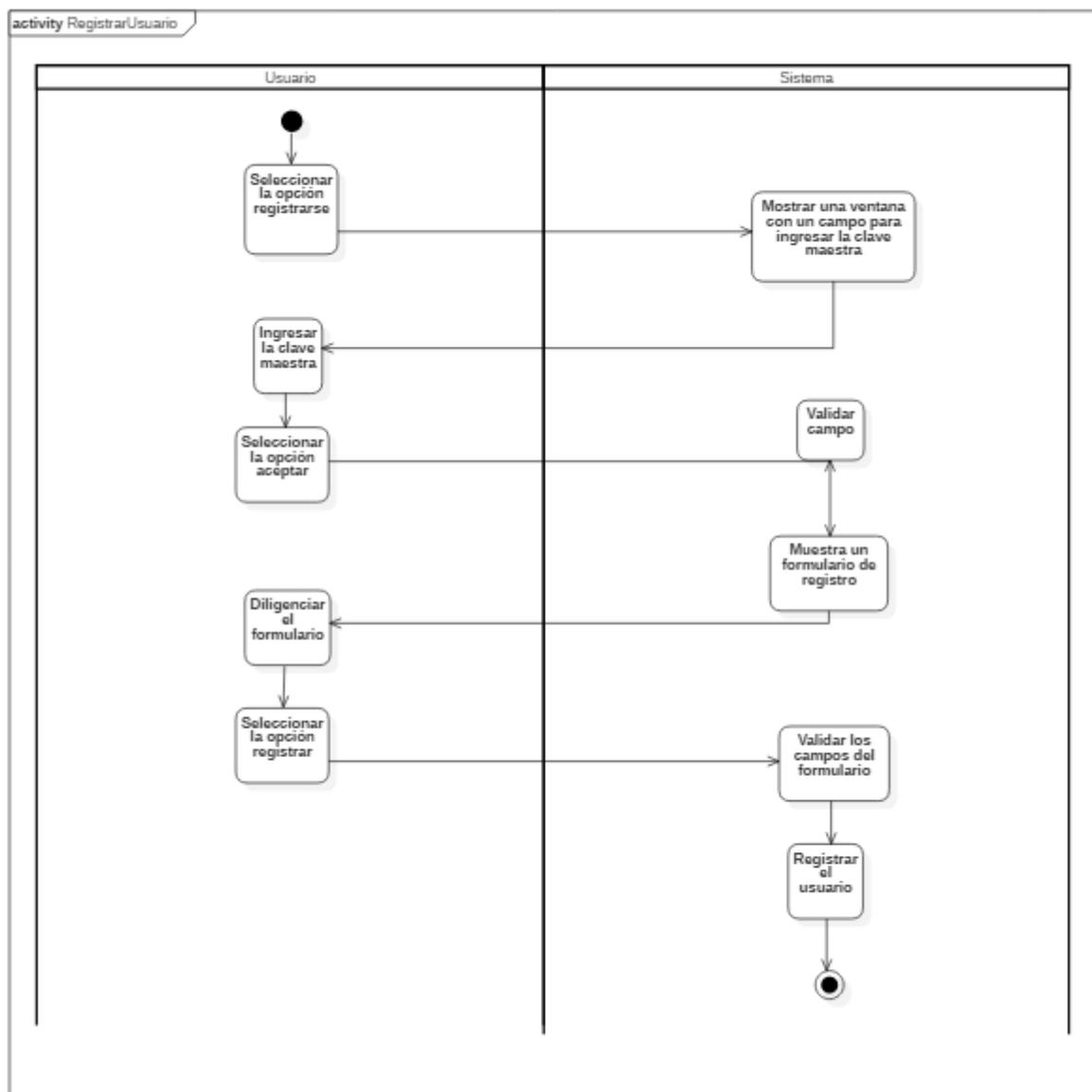


Figura 7. Registrar usuario. Autores del proyecto

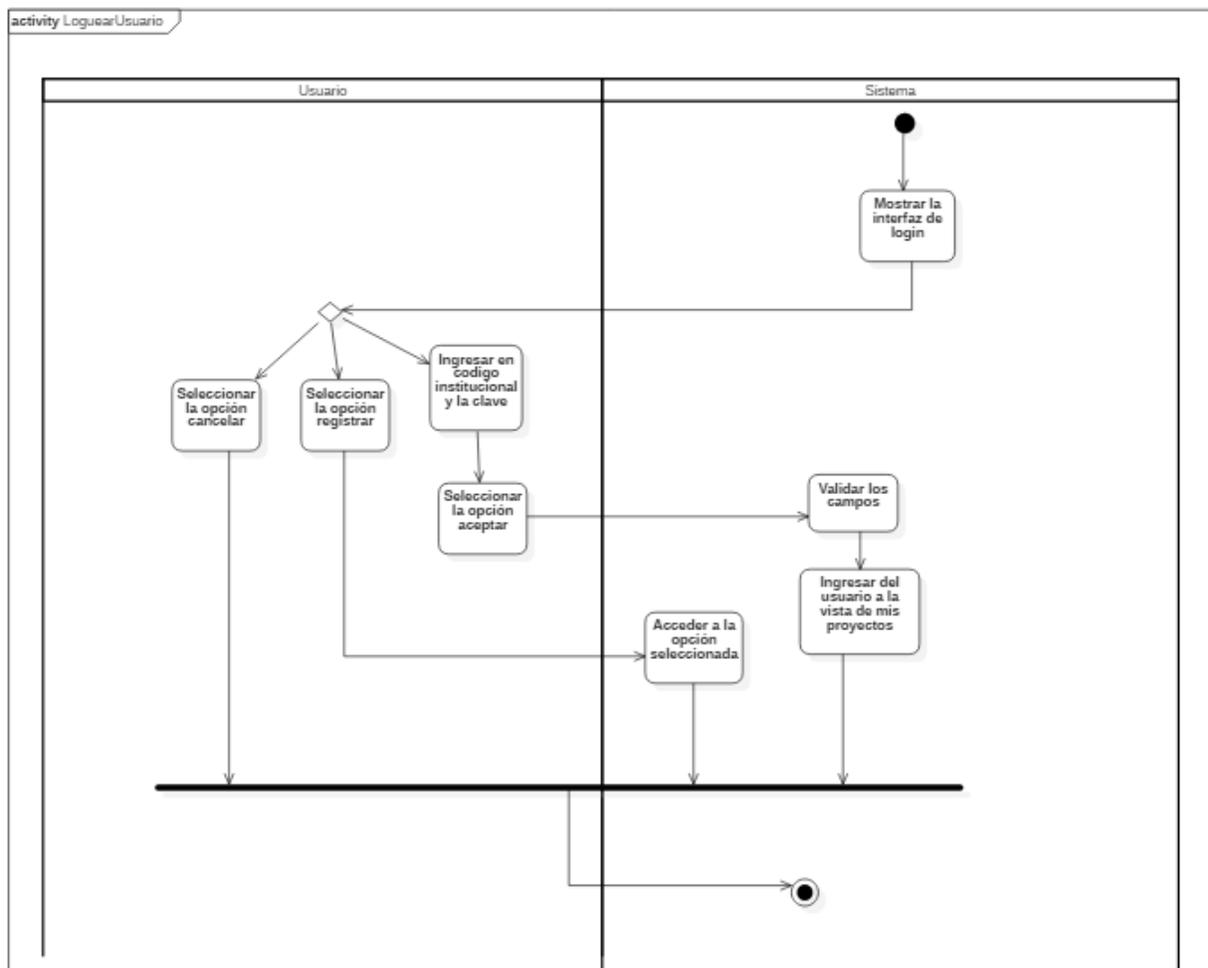


Figura 8. Loguear usuario. Autores del proyecto

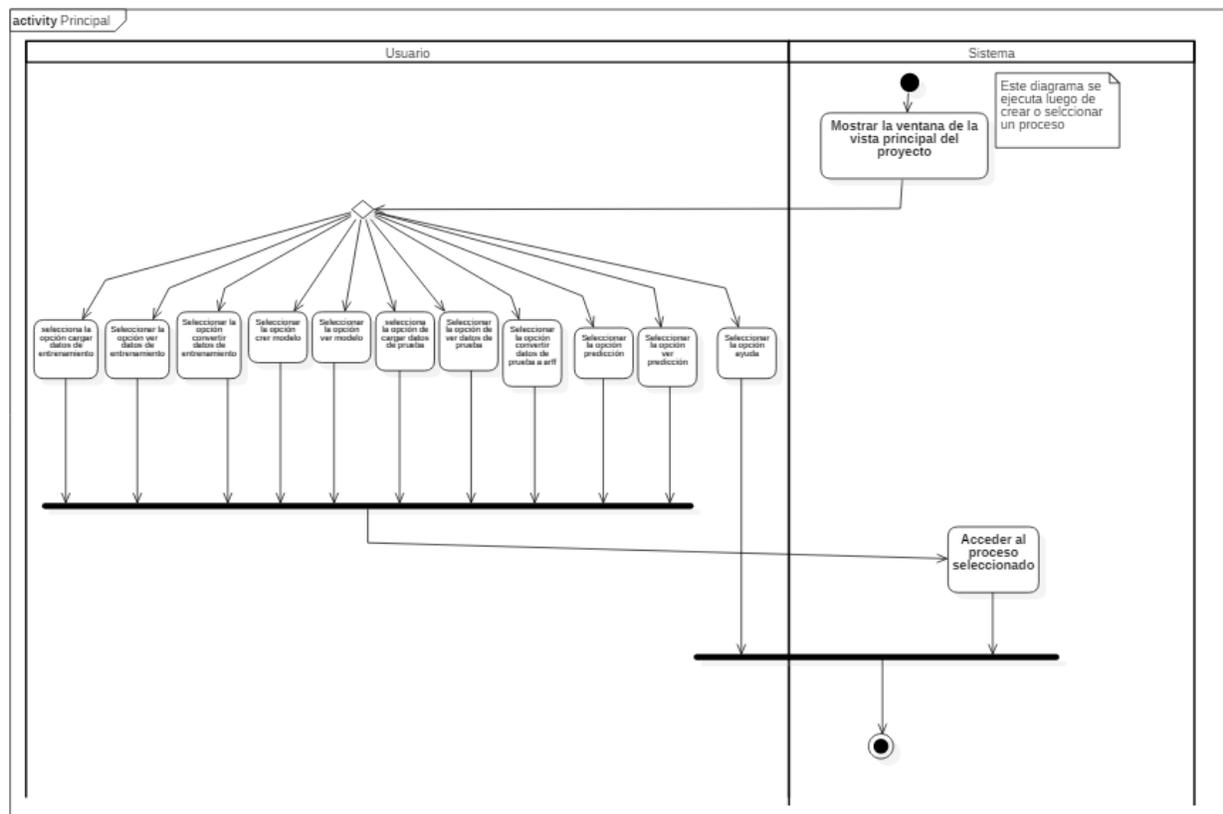


Figura 9. Gestionar principal. Autores del proyecto

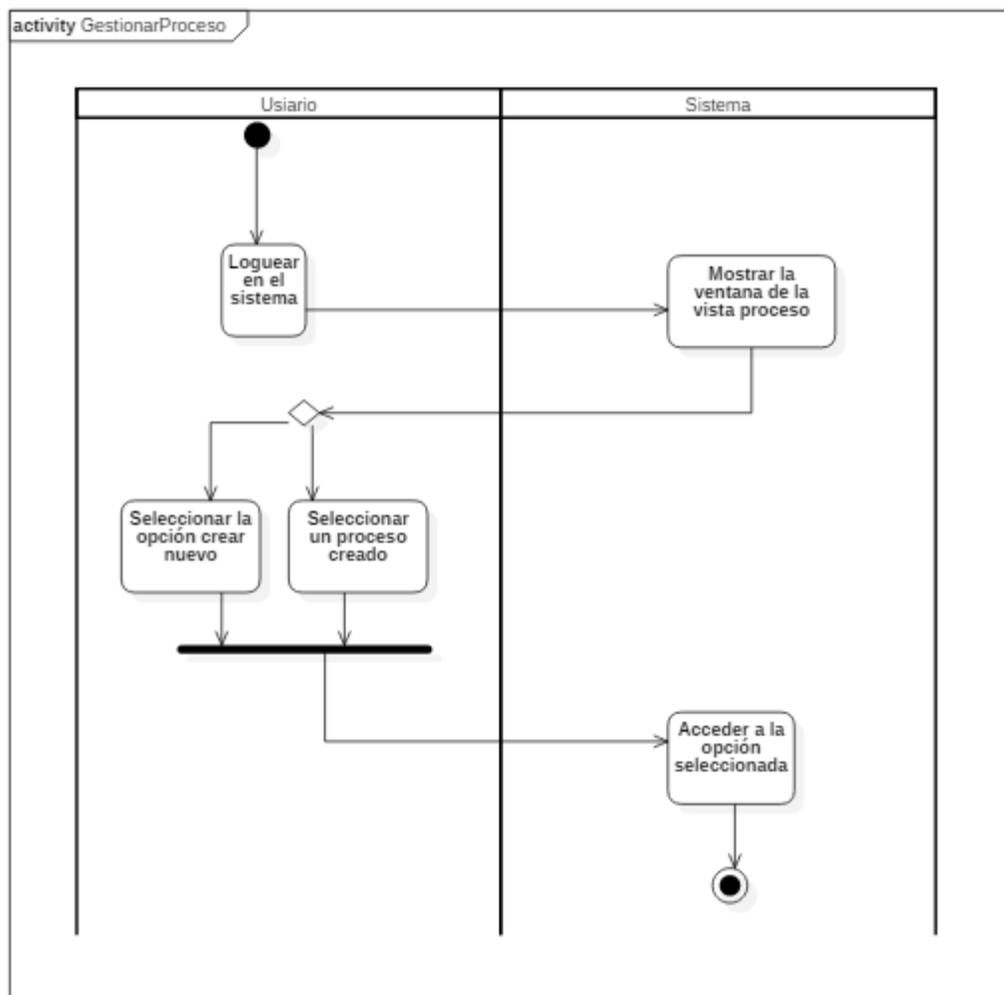


Figura 10. Gestionar proceso. Autores del proyecto

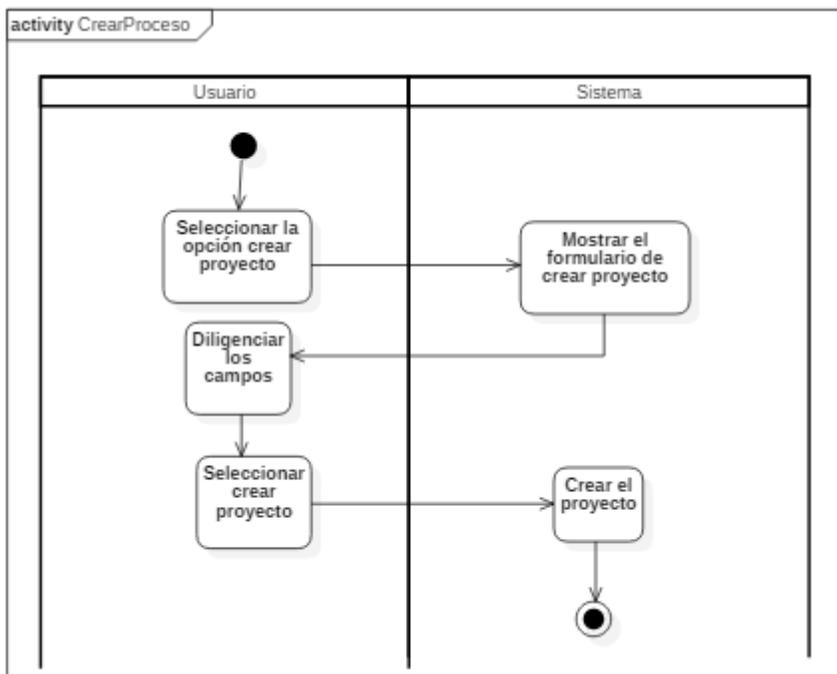


Figura 11. Gestionar información del proyecto: Crear proceso. Autores del proyecto

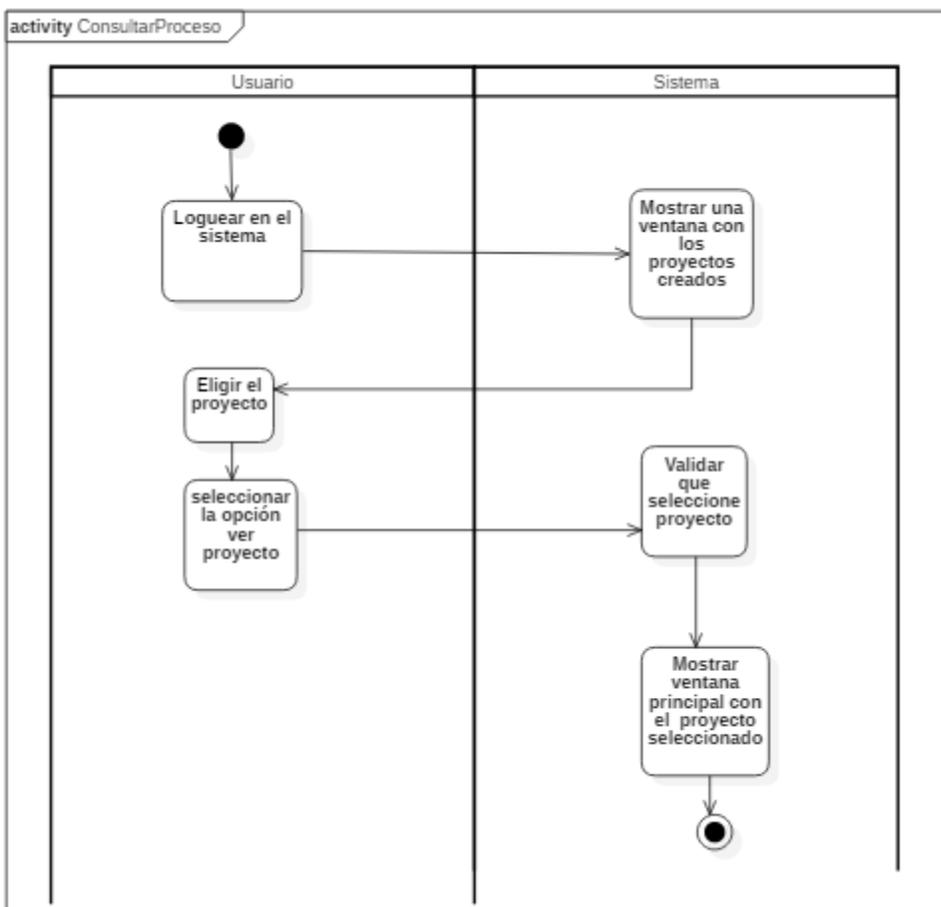


Figura 12. Gestionar información del proyecto: Consultar proyecto. Autores del proyecto

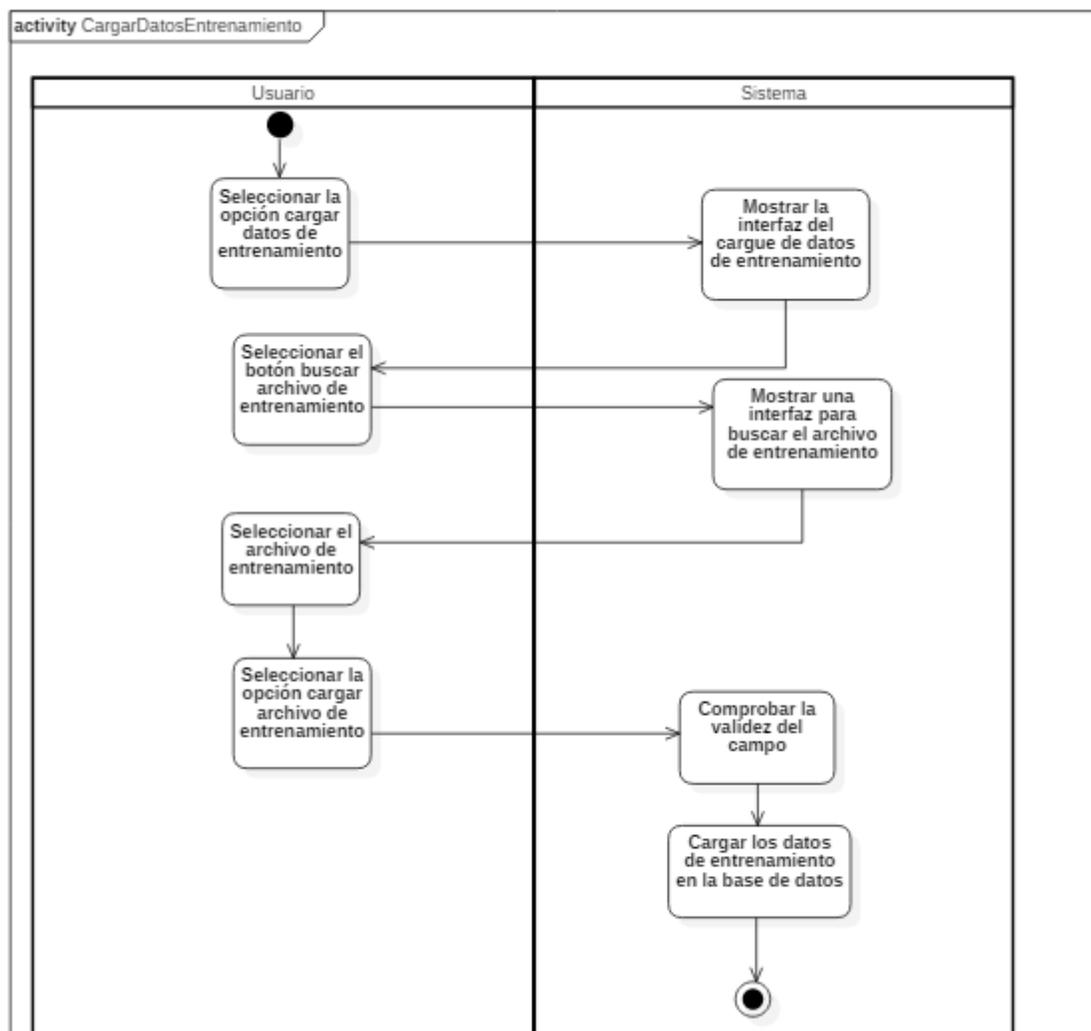


Figura 13. Gestionar datos entrenamiento: Cargar datos de entrenamiento. Autores del proyecto

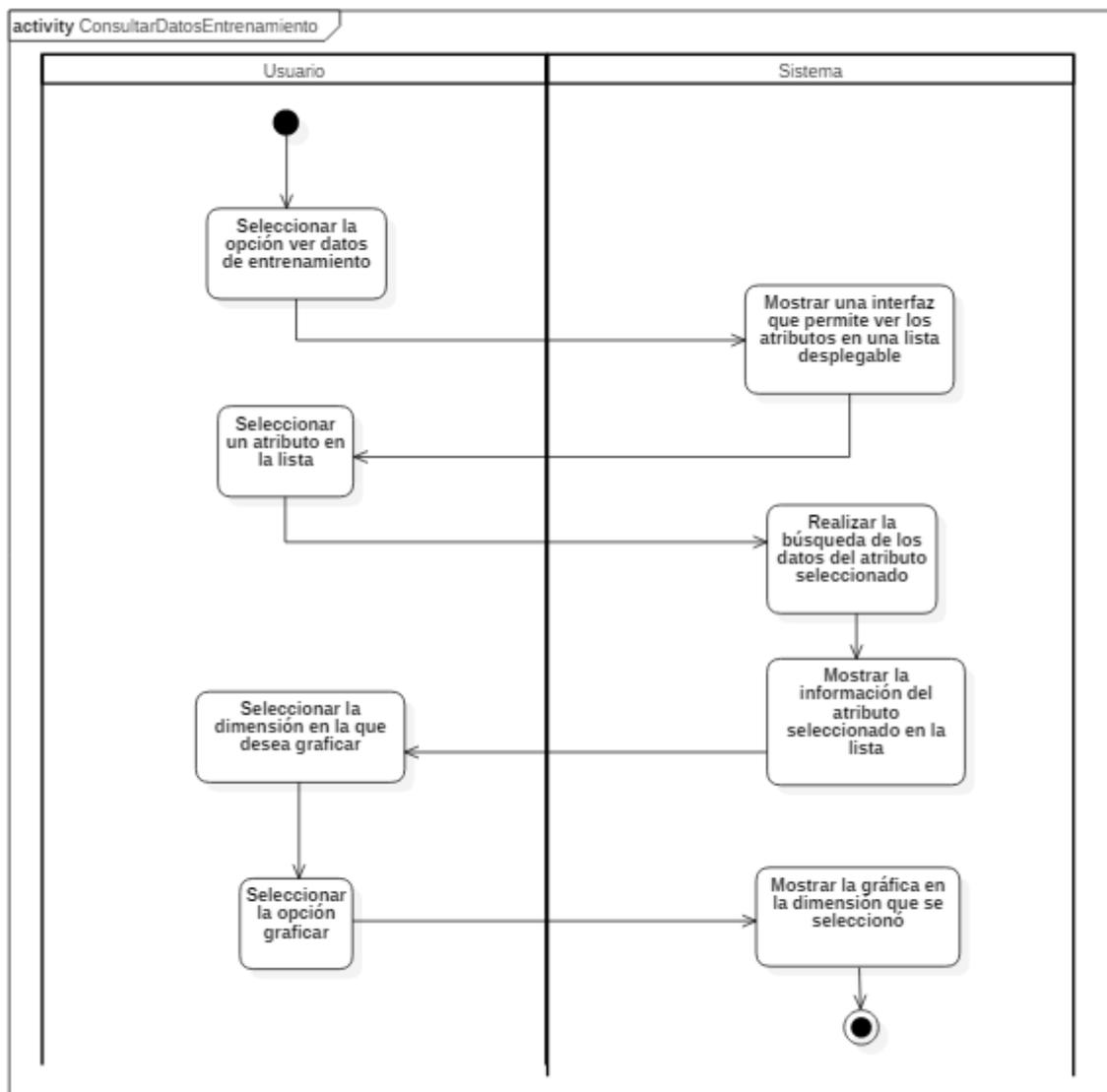


Figura 14. Gestionar datos entrenamiento: Consultar datos de entrenamiento.

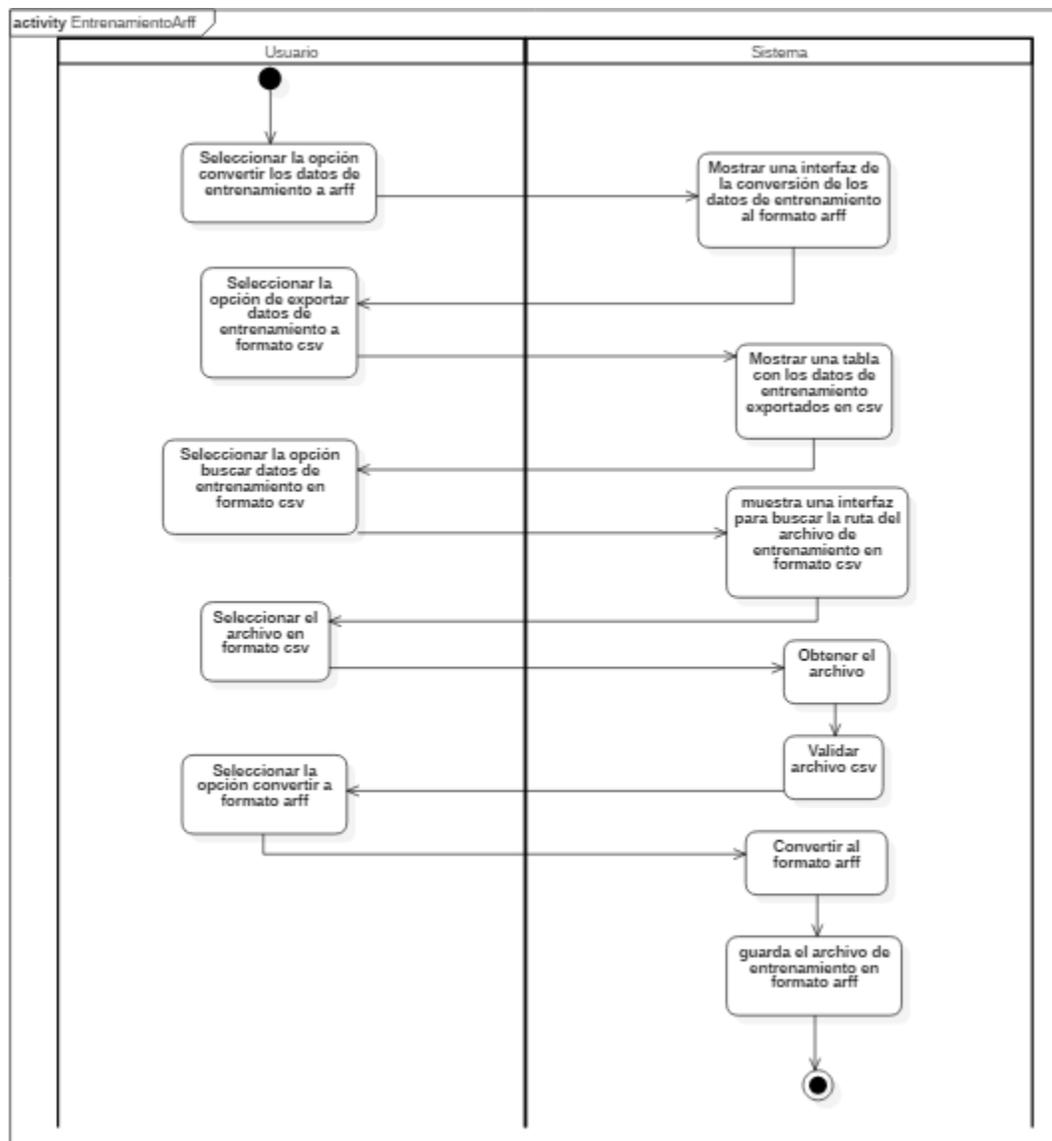


Figura 15. Gestionar minería: Conversión de los datos de entrenamiento al formato arff.

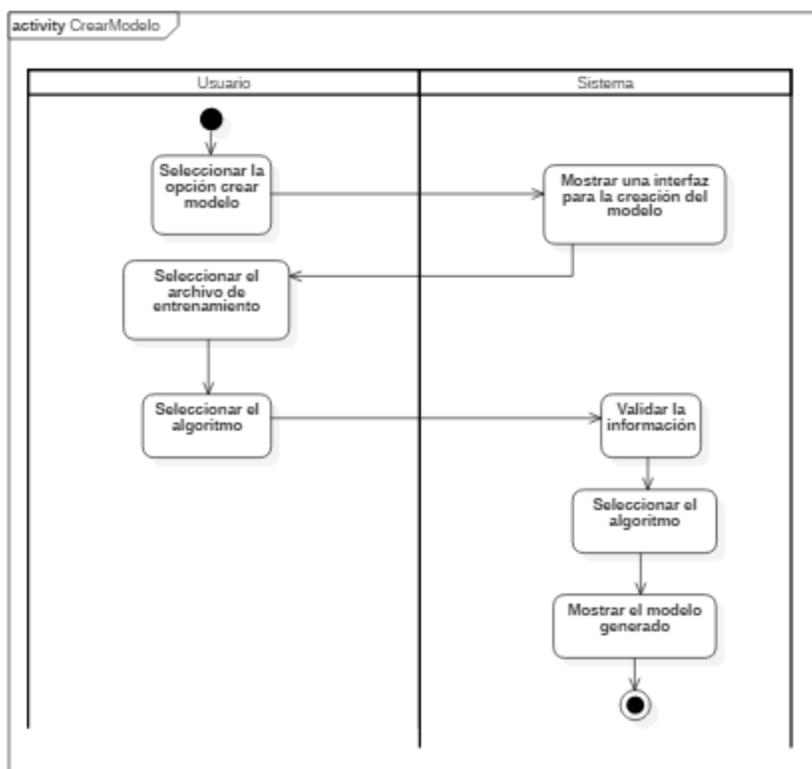


Figura 16. Gestionar minería: Crear modelo. Autores del proyecto

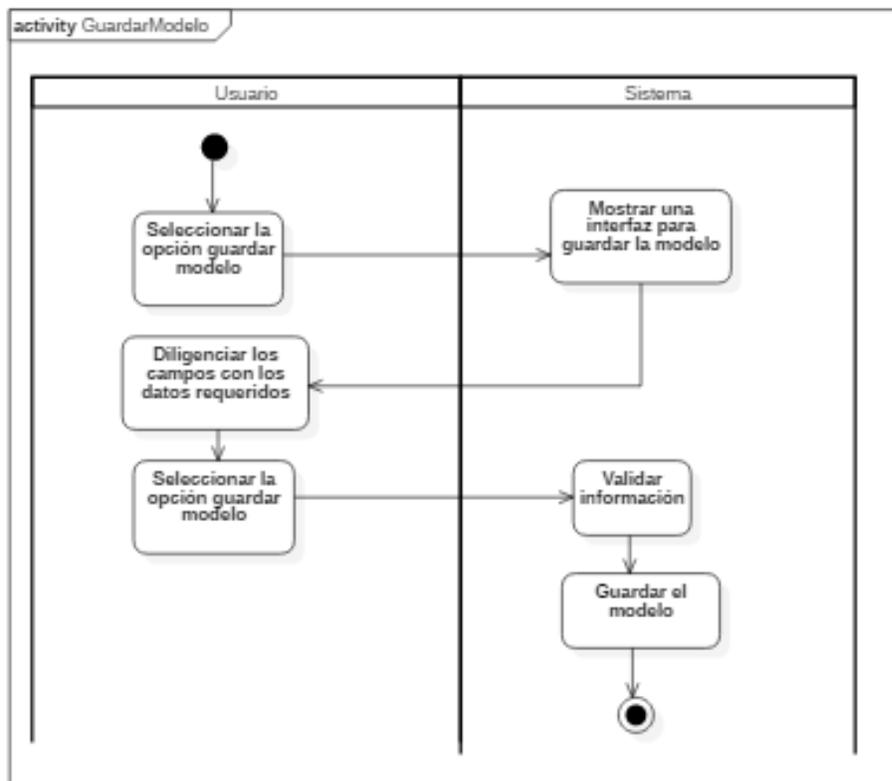


Figura 17. Gestionar minería: Guardar modelo. Autores del proyecto

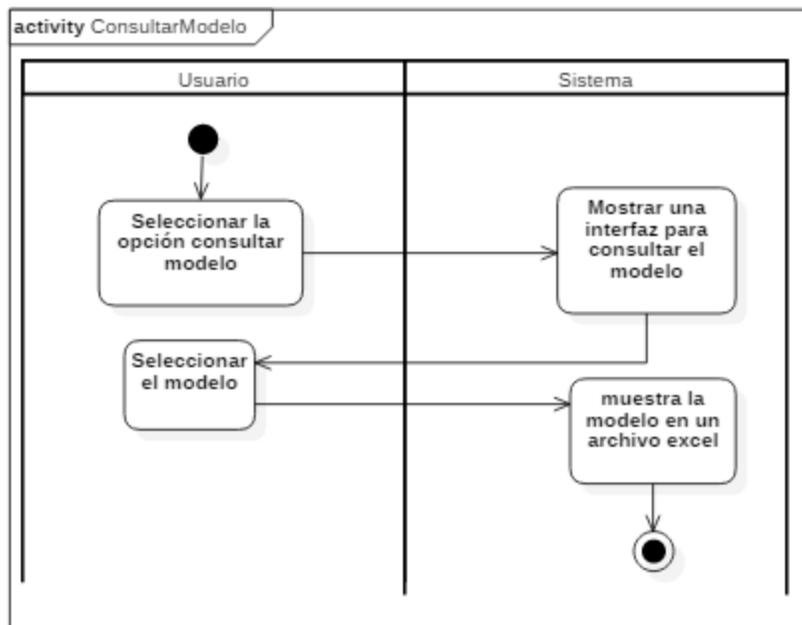


Figura 18. Gestionar minería: Consultar modelo. Autores del proyecto

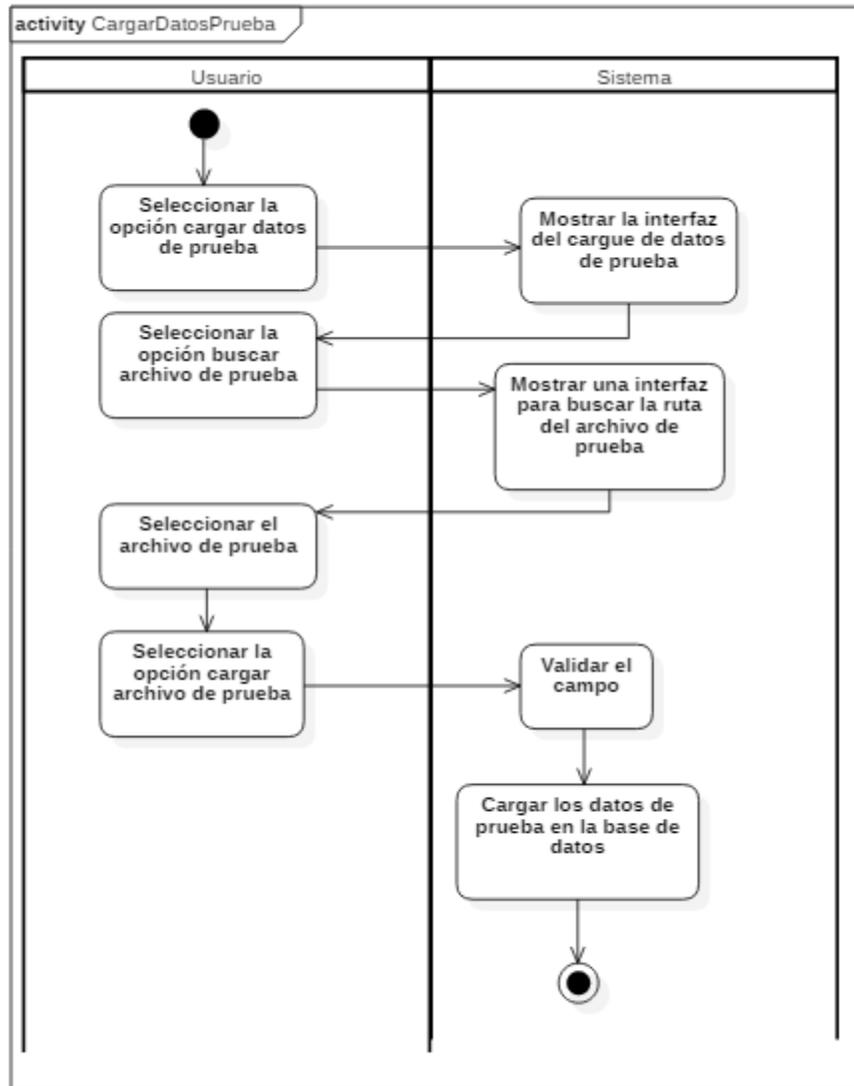


Figura 19. Gestionar datos prueba: Cargar datos de prueba. Autores del proyecto

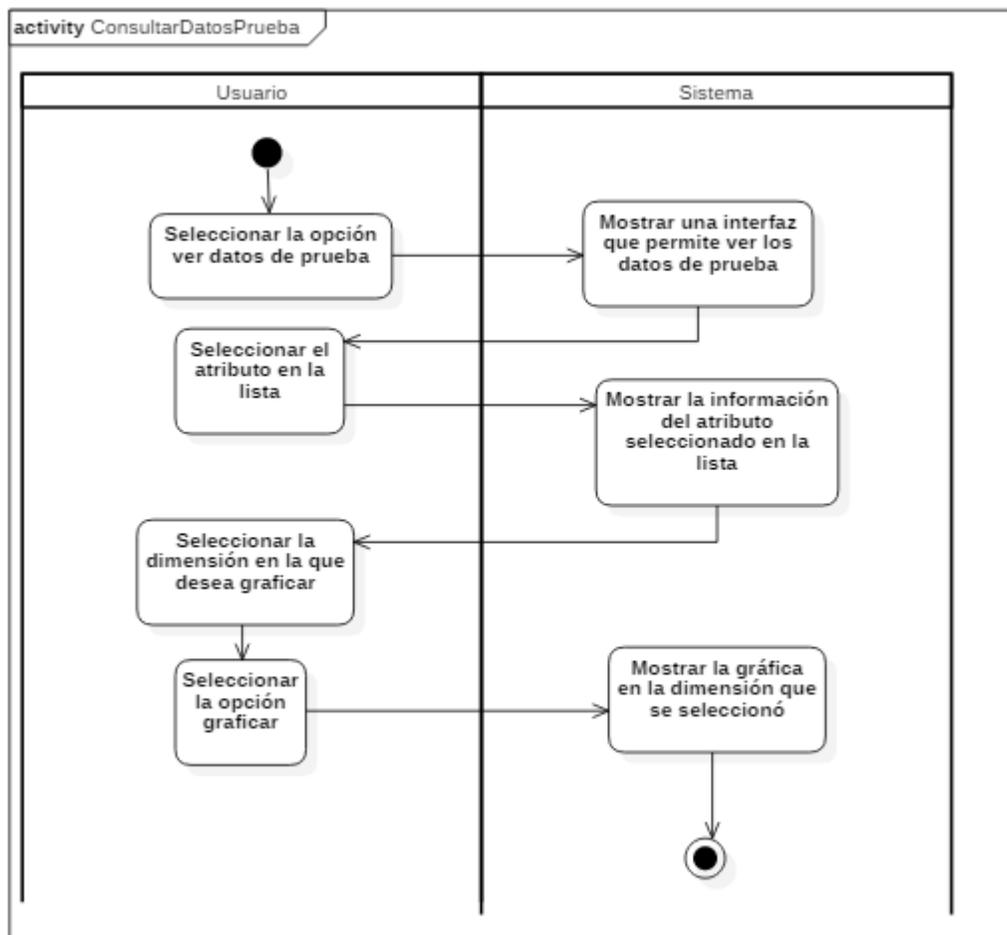


Figura 20. Gestionar datos prueba: Consultar datos de prueba. Autores del proyecto

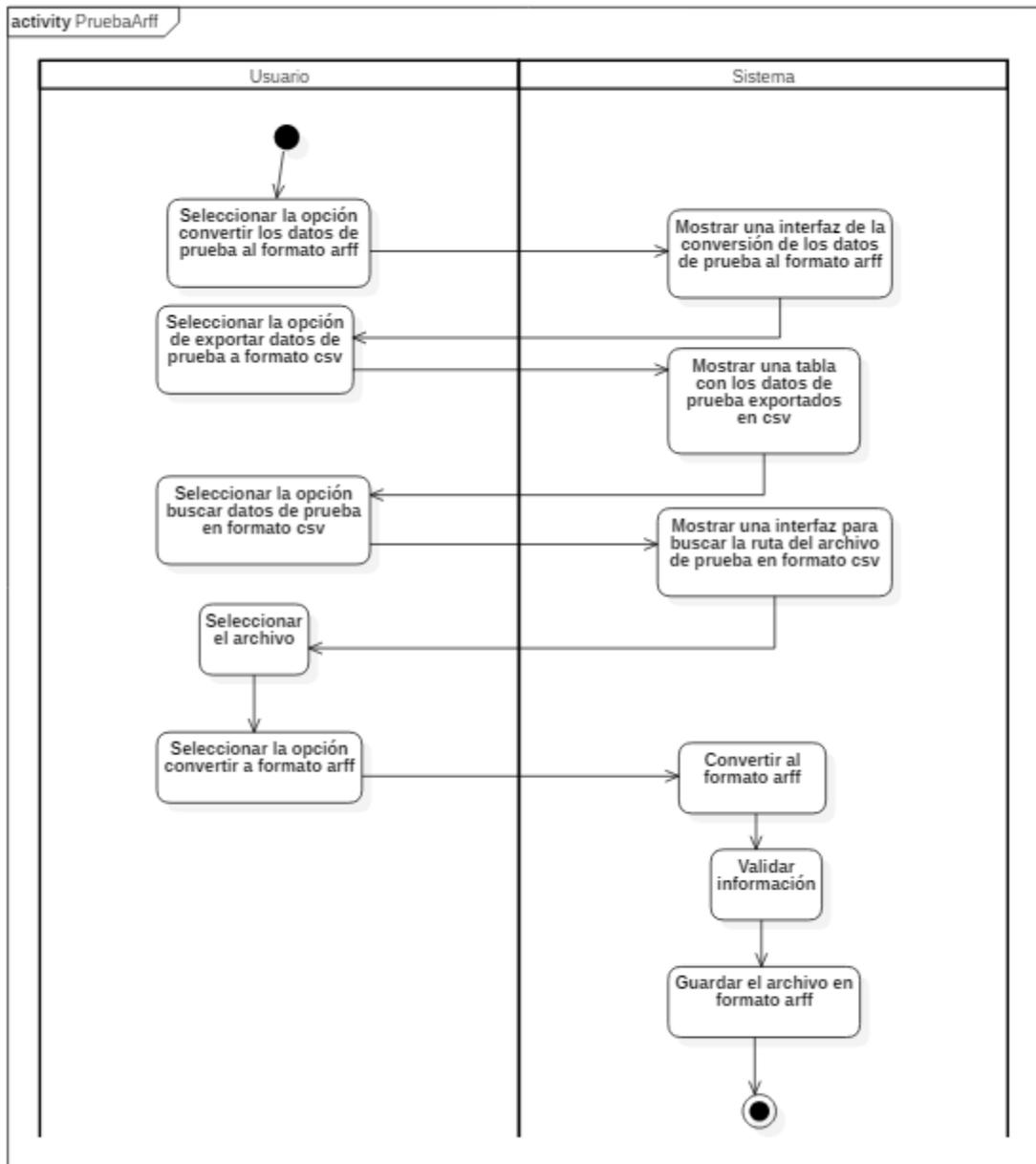


Figura 21. Gestionar predicción: Conversión de los datos de prueba al formato arff. Autores del proyecto

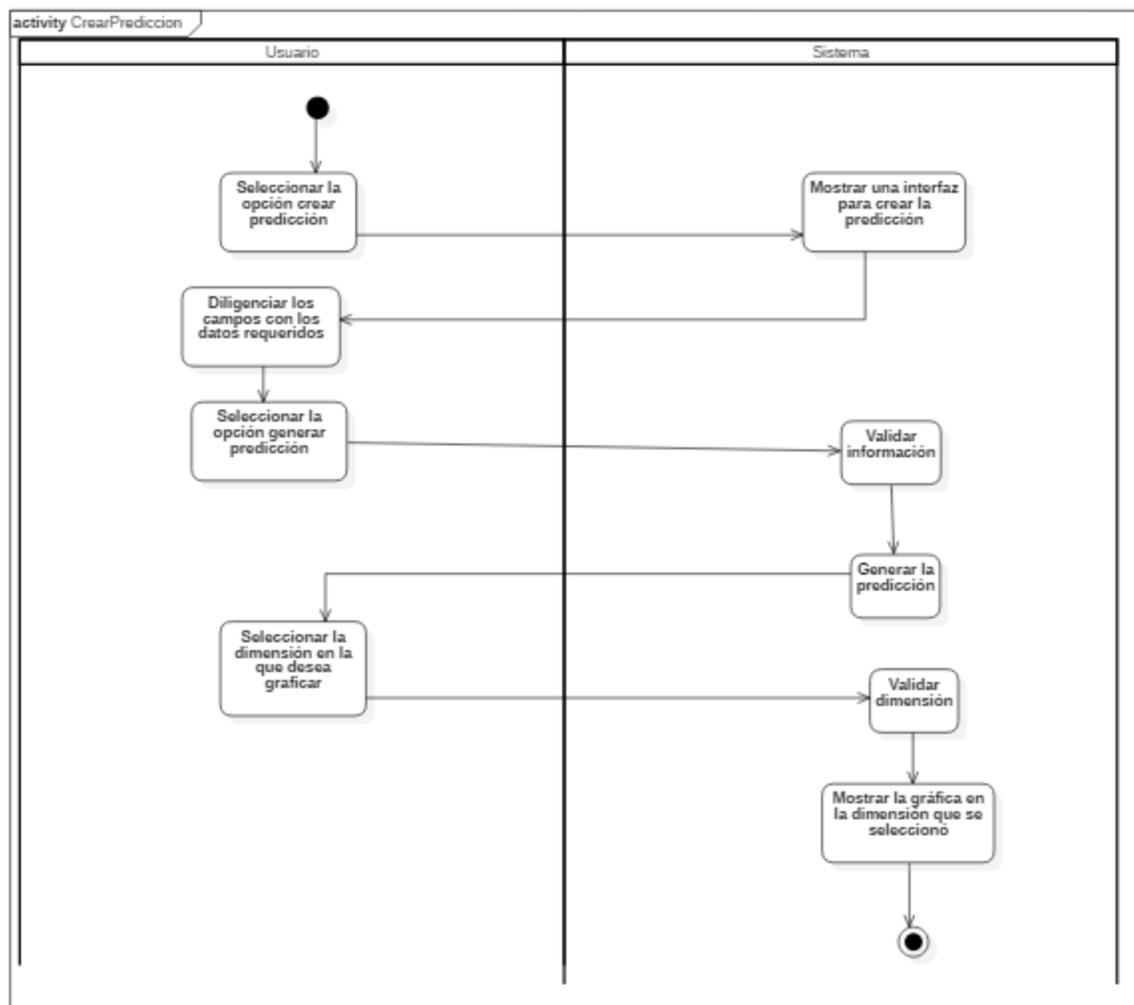


Figura 22. Gestionar predicción: Crear predicción. Autores del proyecto

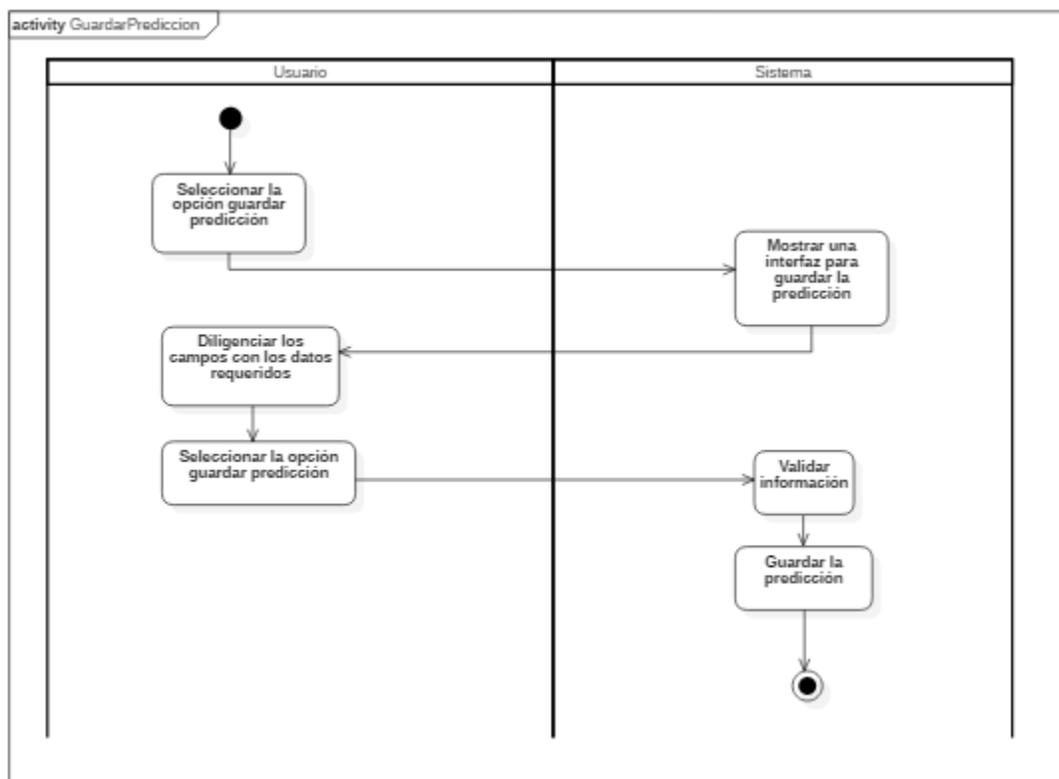


Figura 23. Gestionar predicción: Guardar predicción. Autores del proyecto

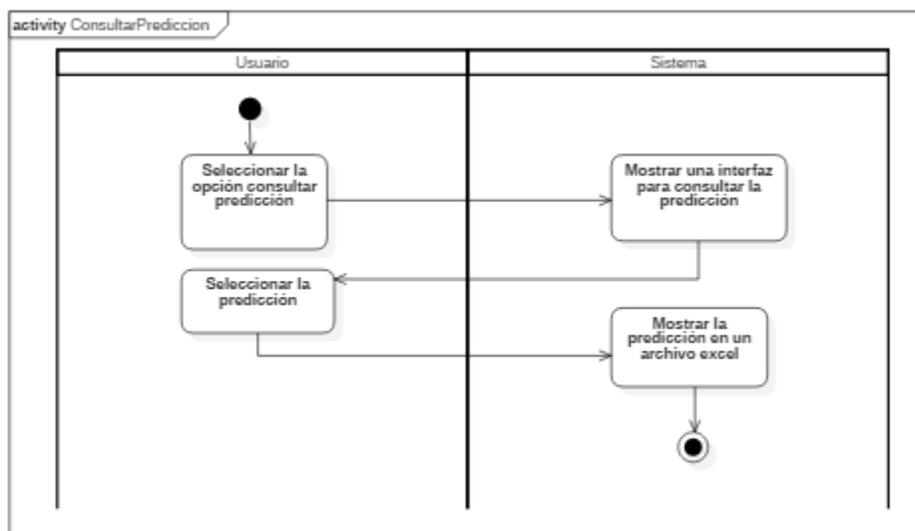


Figura 24. Gestionar predicción: Consultar predicción. Autores del proyecto

4.3.2.3 Fase de construcción iteración. En esta fase se desarrollan los diagramas del sistema y el producto final logrando los objetivos planteados.

4.3.2.3.1 Modelo relacional.

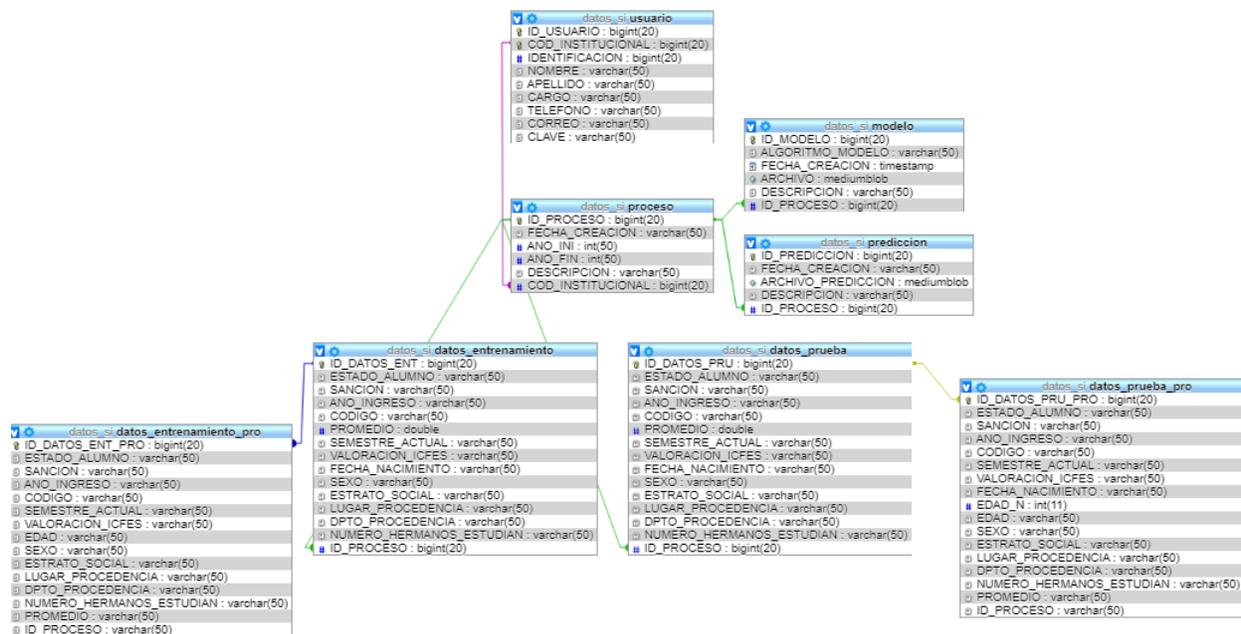


Figura 25. Modelo relacional. Autores del proyecto

4.3.2.3.2 Arquitectura en capas. Es la arquitectura con la que está programado el software. Contiene tres capas que son: capa de presentación, capa lógica y la capa de la persistencia.

Capa de presentación: es la capa que interactúa directamente con el usuario y es la que recibe y emite las notificaciones de cambios en el sistema.

Capa de lógica de negocios: esta capa se encarga de proteger el acceso directo a la información y lleva un control sobre la información que se transita desde la capara de la persistencia hasta la capa de presentación y viceversa.

Capa de persistencia: Es responsable del almacenamiento de los datos, en esta se encuentran capas incluso más detalladas que van desde fachada hasta DAO u estrategia dependiendo de la necesidad en la programación (Daniel, 2001).

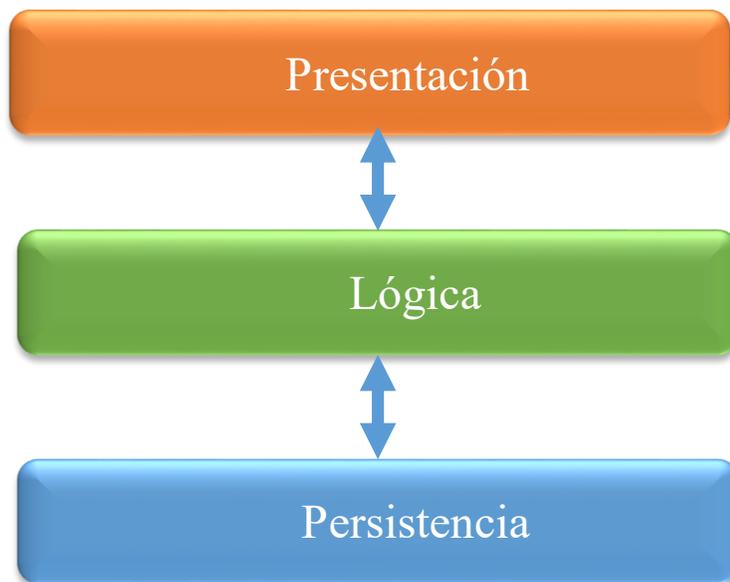


Figura 26. Arquitectura por capas. Autores del proyecto

4.3.2.3.3 Diagrama de paquetes.

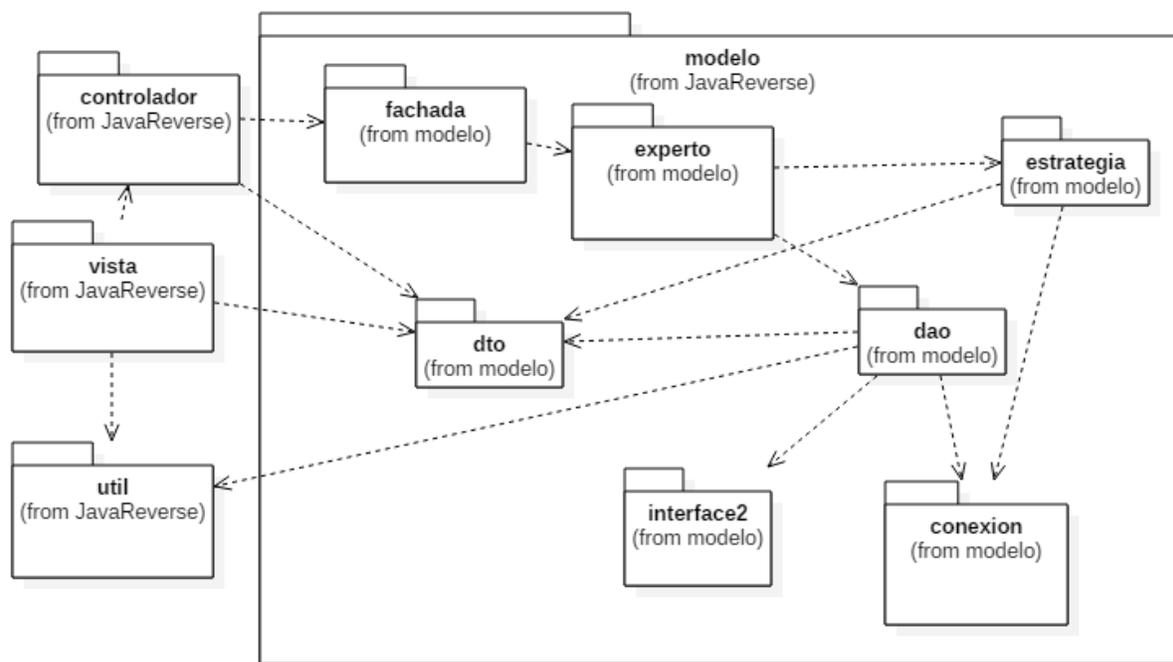


Figura 27. Diagrama de paquetes. Autores del proyecto

4.3.2.3.4 Diagramas de clases.

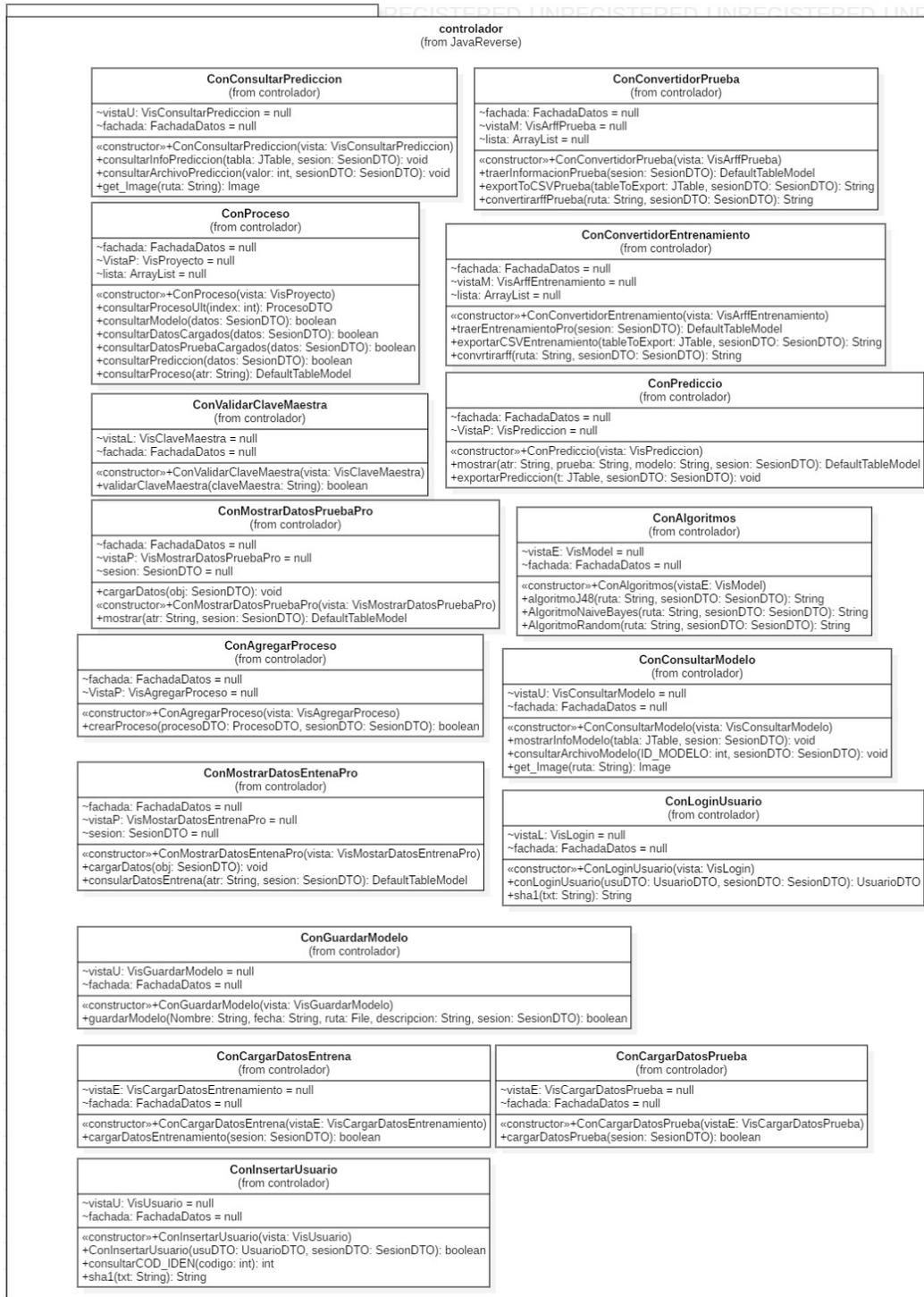


Figura 28. Diagrama de clase: controlador. Autores del proyecto

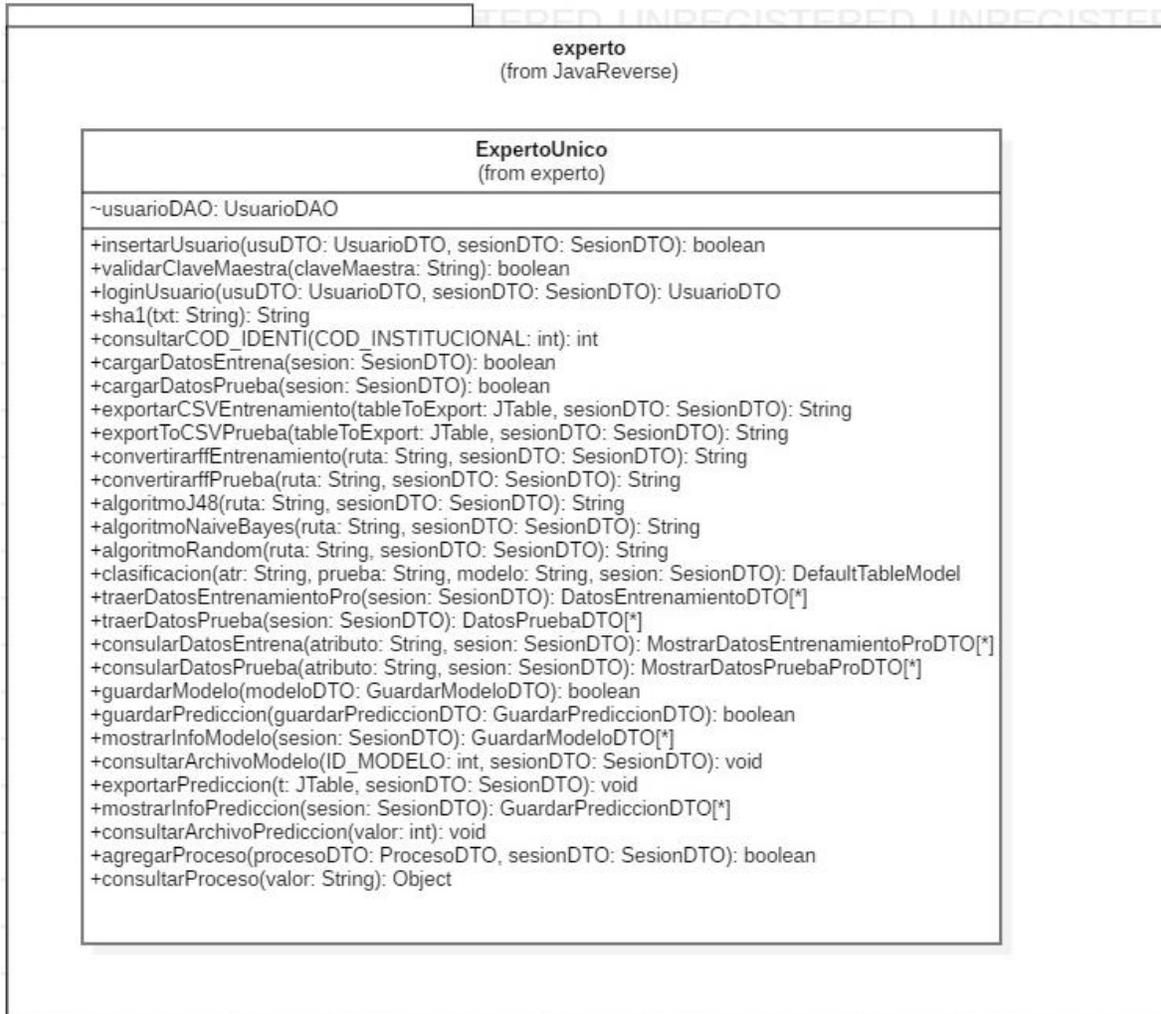


Figura 29. Diagrama de clase: experto. Autores del proyecto

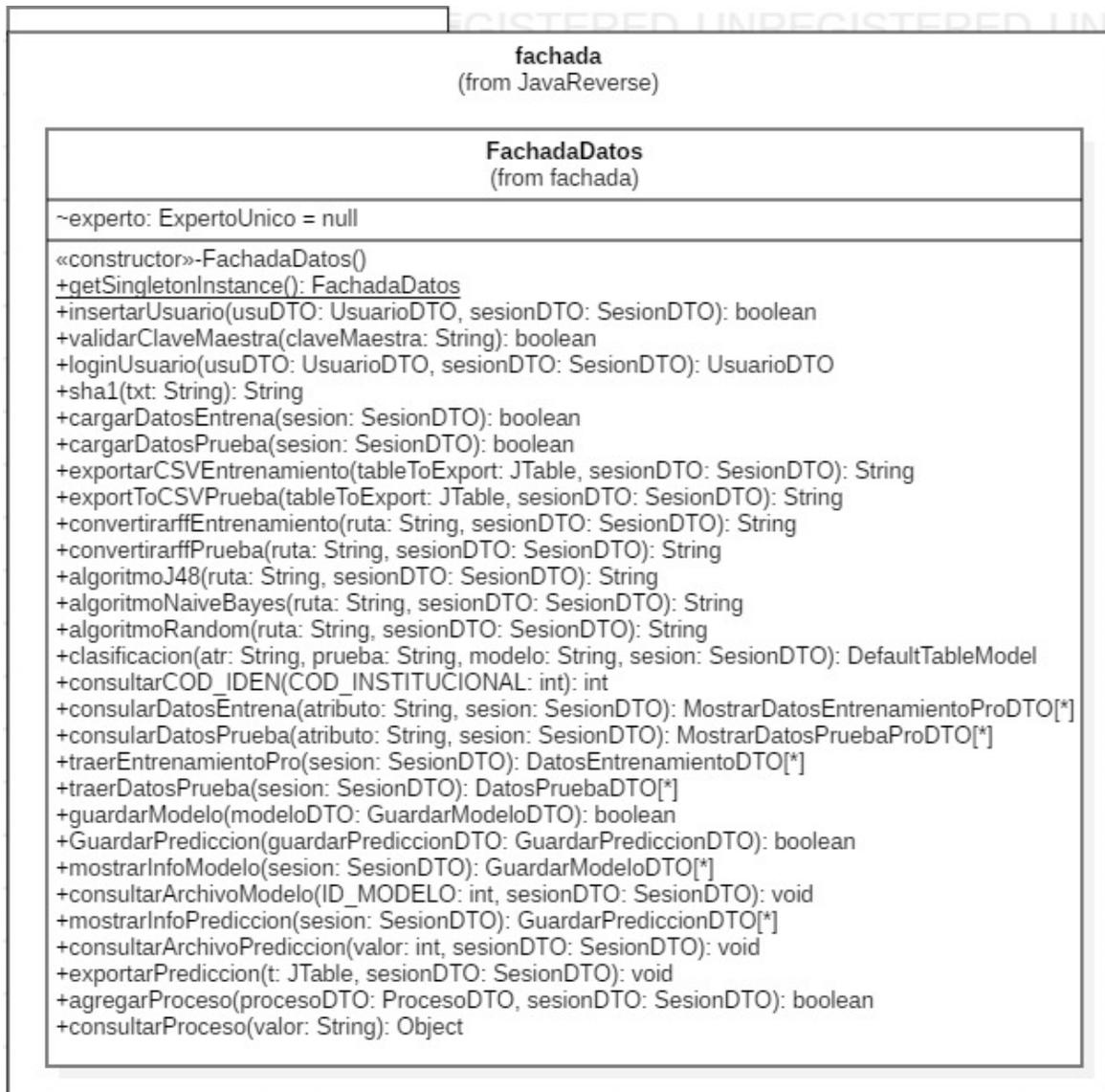


Figura 30. Diagrama de clase: fachada. Autores del proyecto

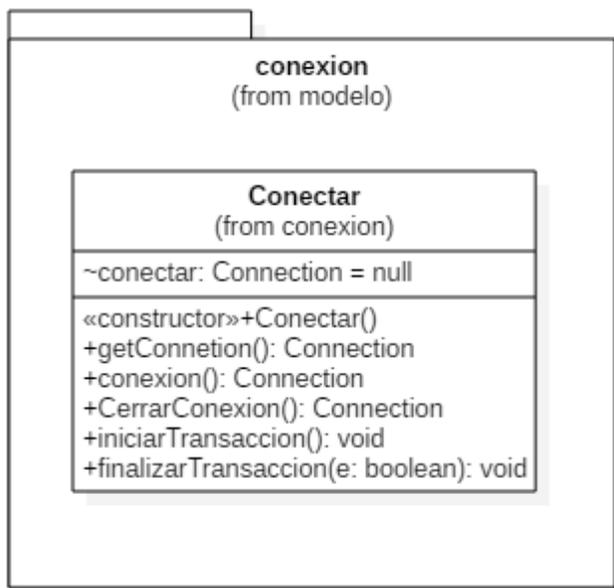


Figura 31. Diagrama de clase: conexión. Autores del proyecto

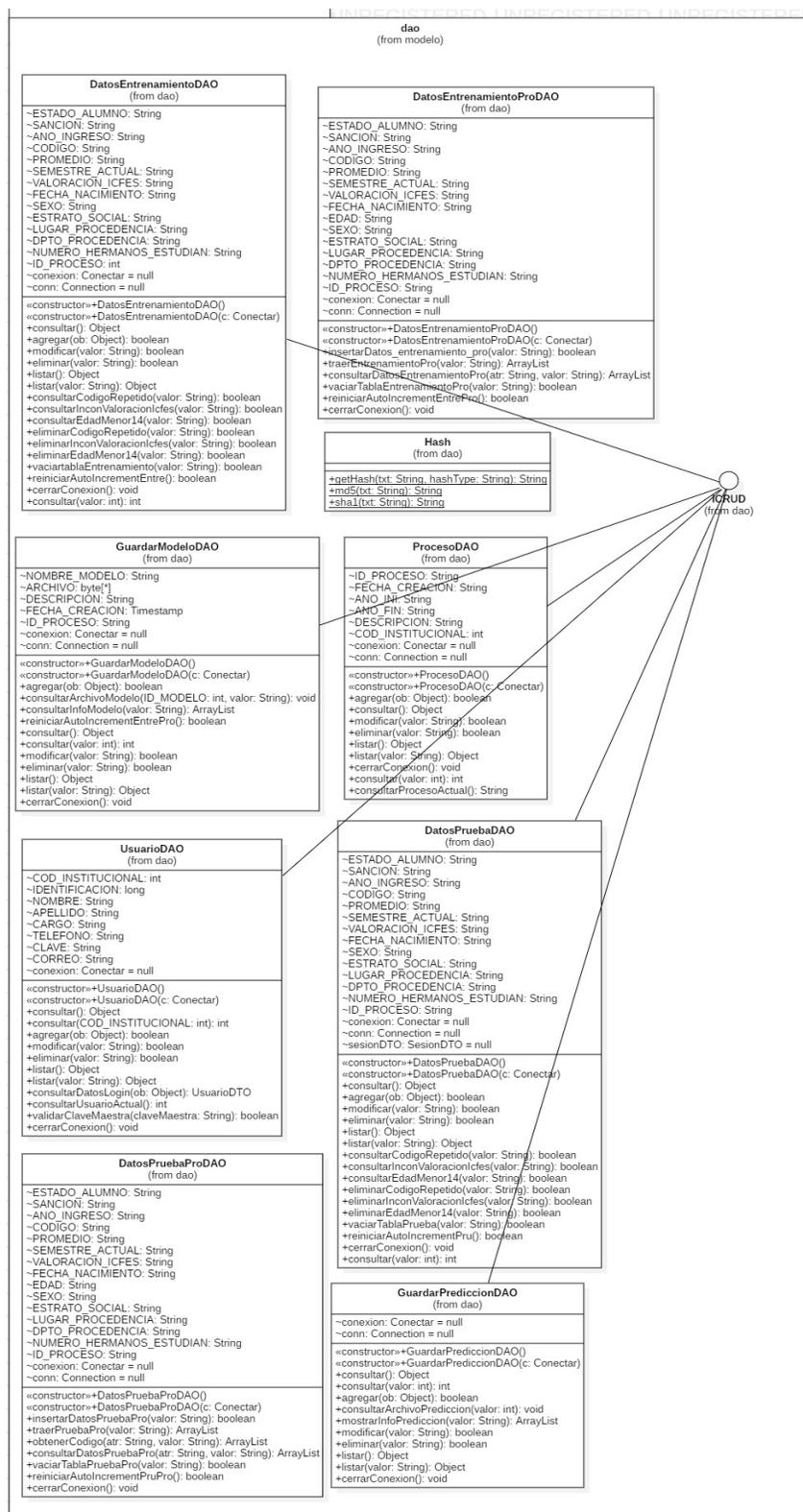


Figura 32. Diagrama de clase: dao. Autores del proyecto

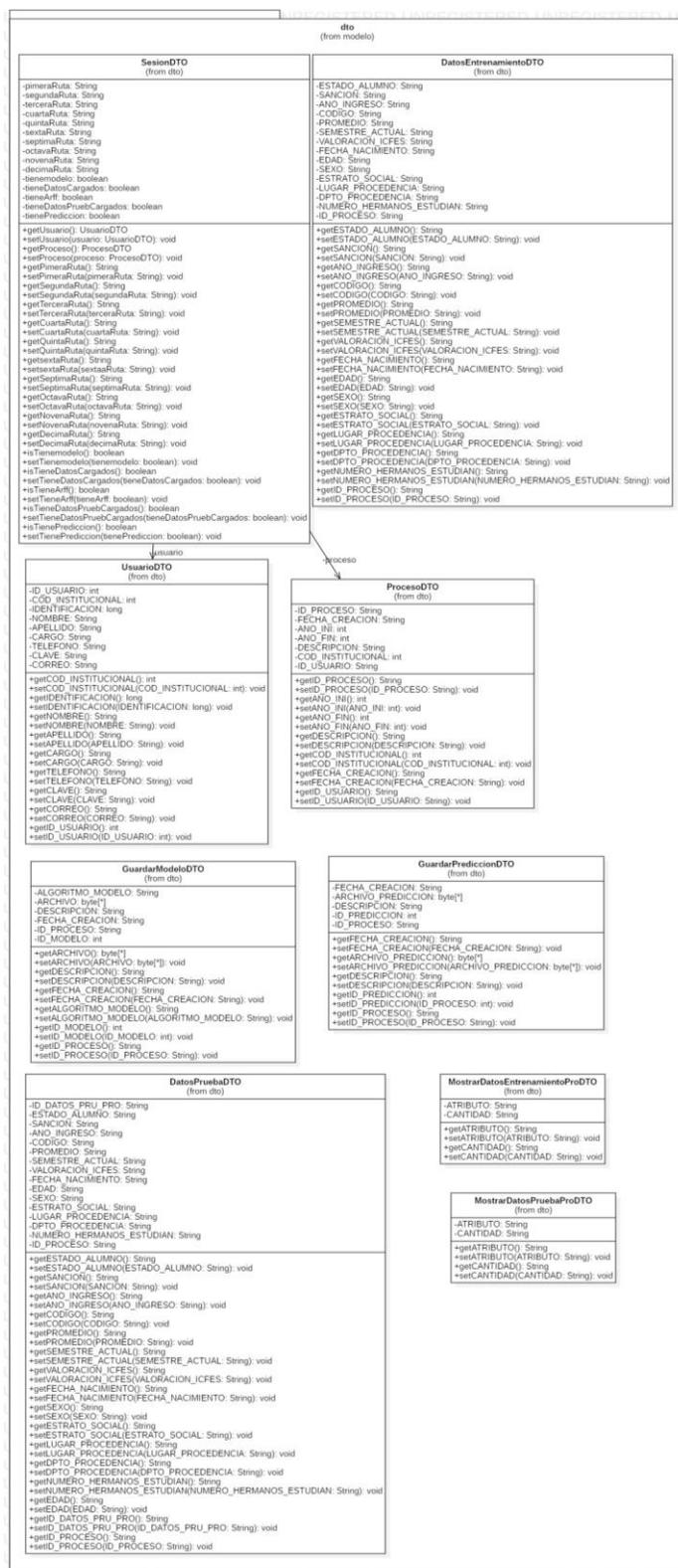


Figura 33. Diagrama de clase: dto. Autores del proyecto

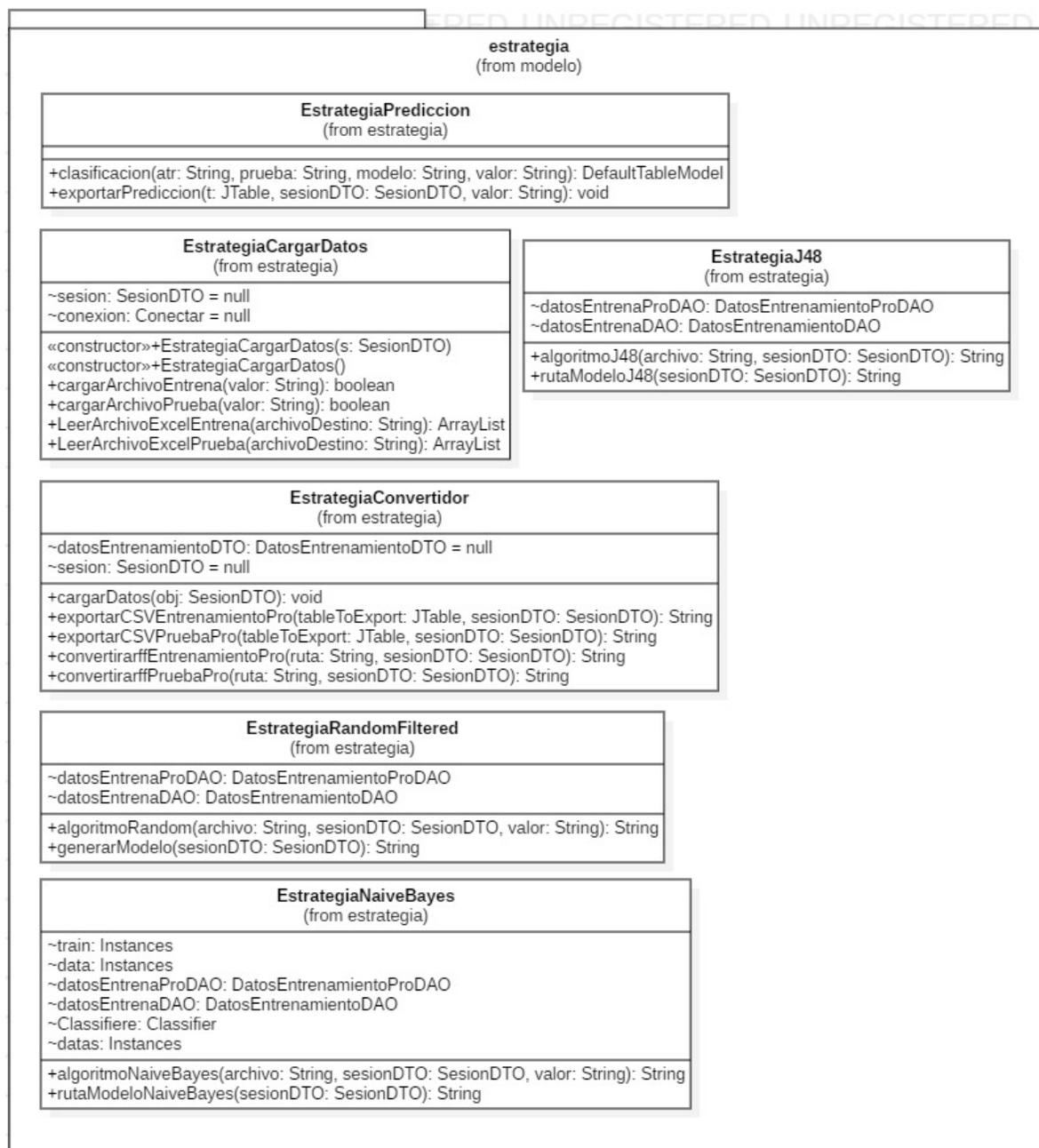


Figura 34. Diagrama de clase: estrategia. Autores del proyecto

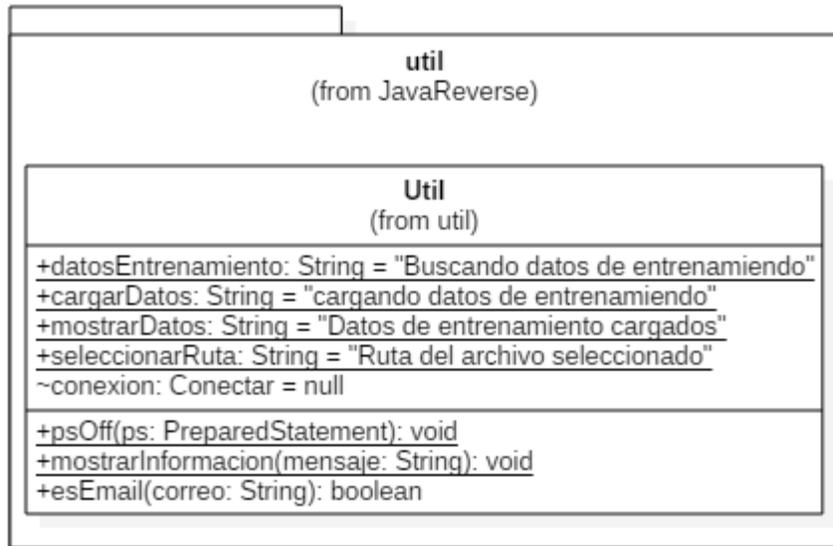


Figura 35. Diagrama de clase: útil. Autores del proyecto

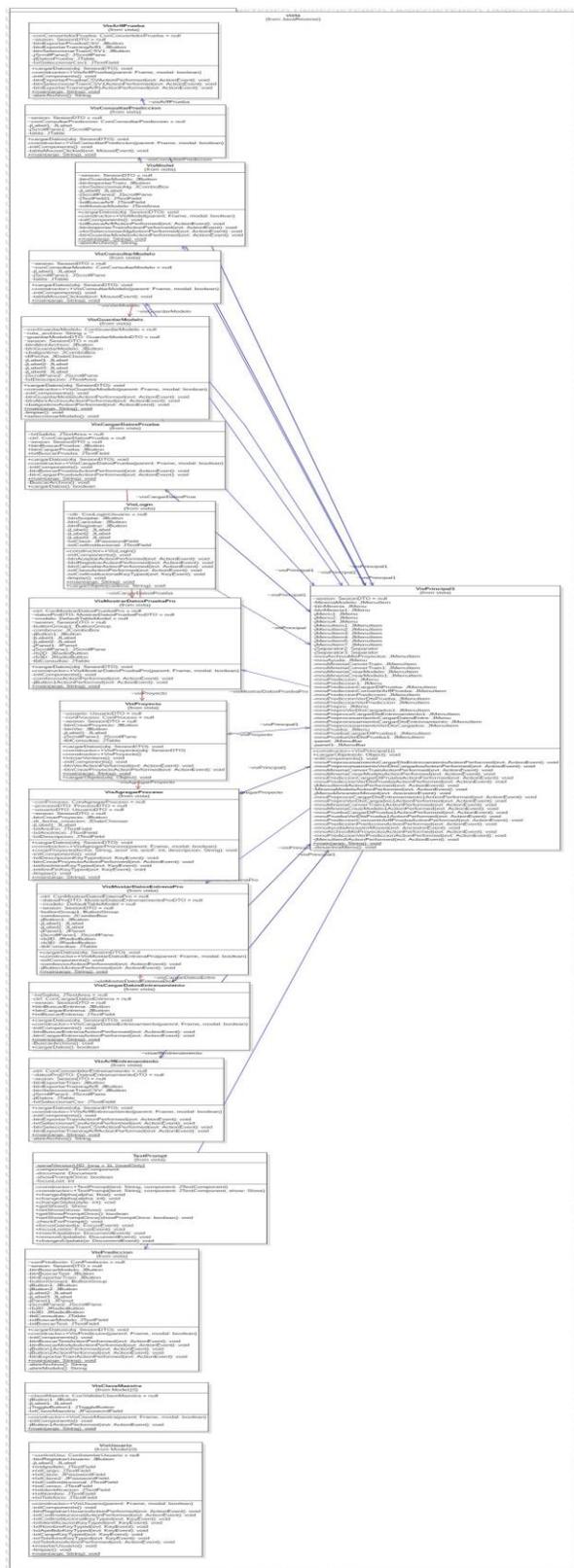


Figura 36. Diagrama de clase: vista. Autores del proyecto

4.3.2.3.5 Diagramas de secuencia.

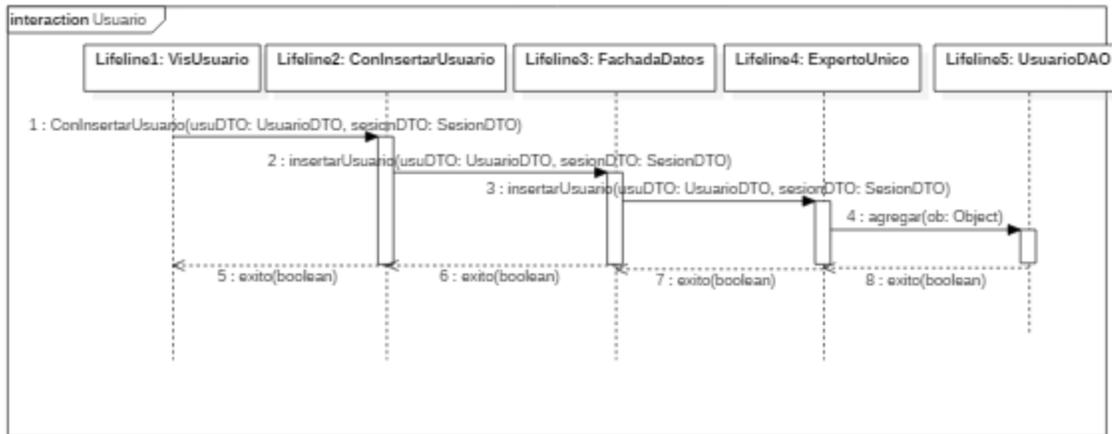


Figura 37. Diagrama de secuencia: Usuario. Autores de proyecto

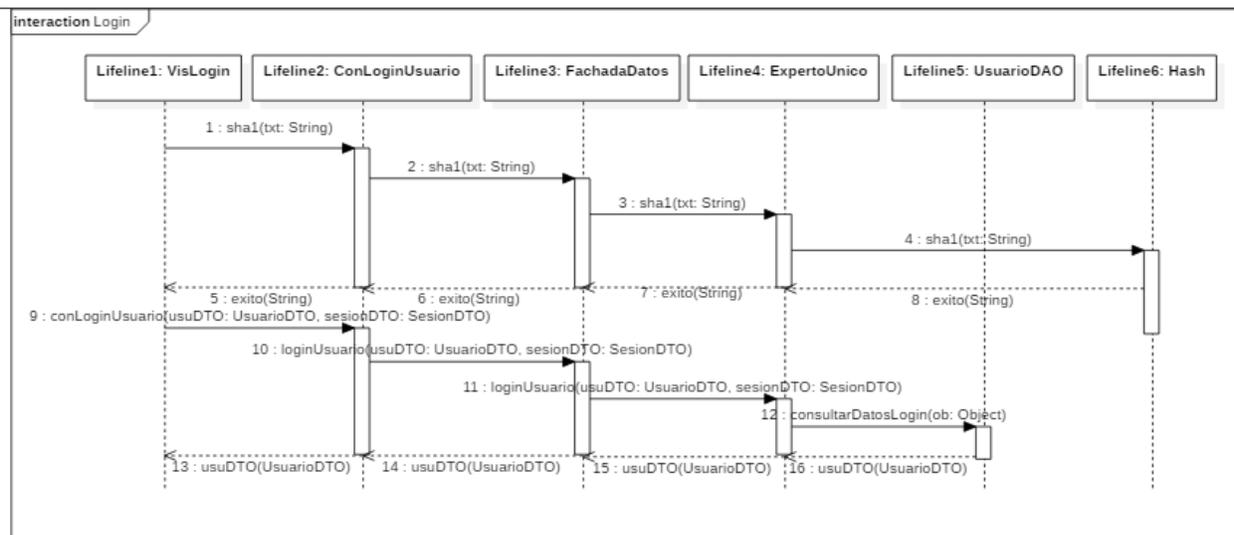


Figura 38. Diagrama de secuencia: Login. Autores de proyecto

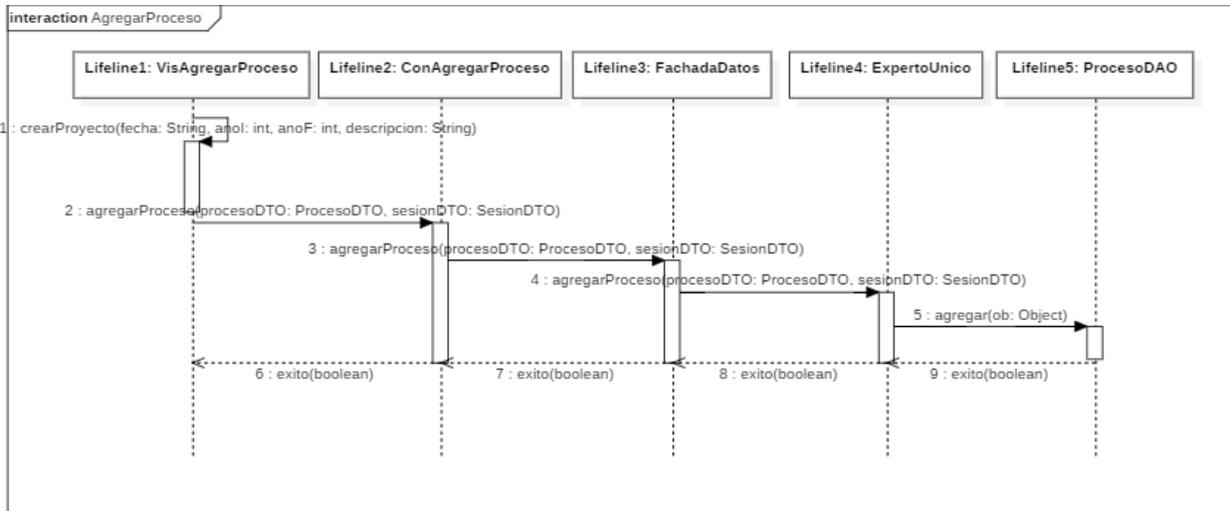


Figura 39. Diagrama de secuencia: Agregar proceso. Autores de proyecto

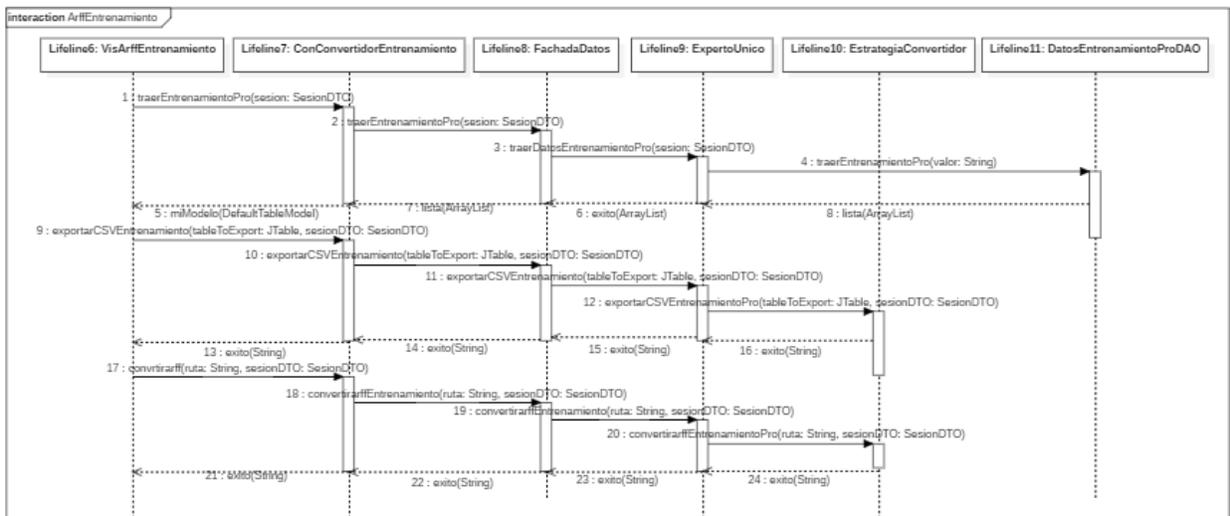


Figura 40. Diagrama de secuencia: Arff entrenamiento. Autores de proyecto

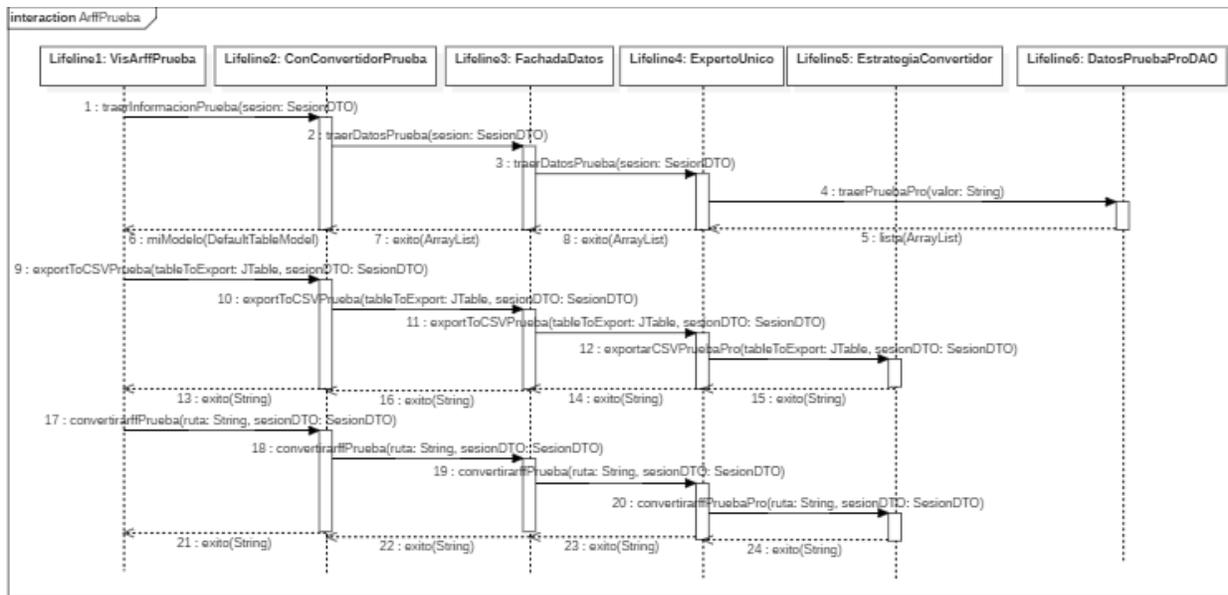


Figura 41. Diagrama de secuencia: Arff prueba. Autores de proyecto

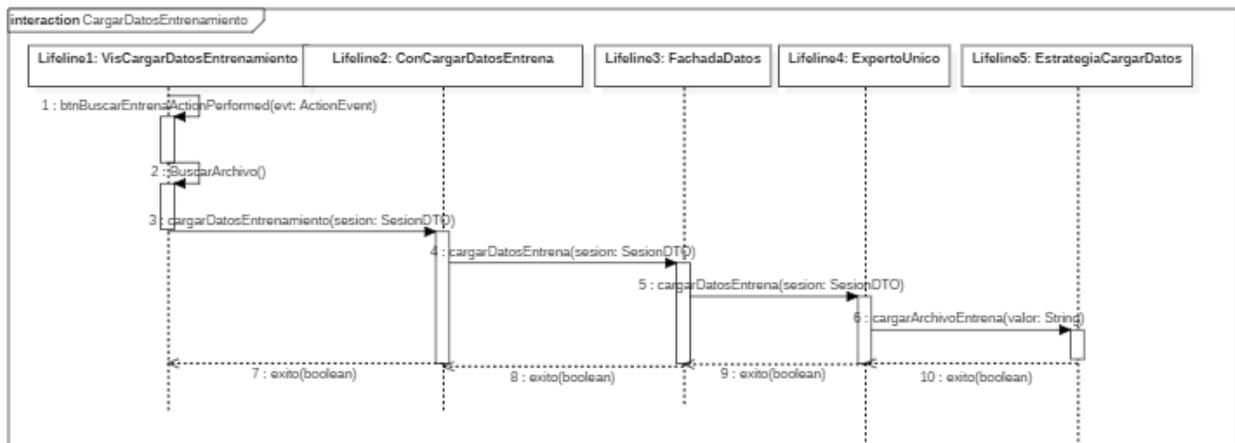


Figura 42. Diagrama de secuencia: Cargar datos de entrenamiento. Autores de proyecto

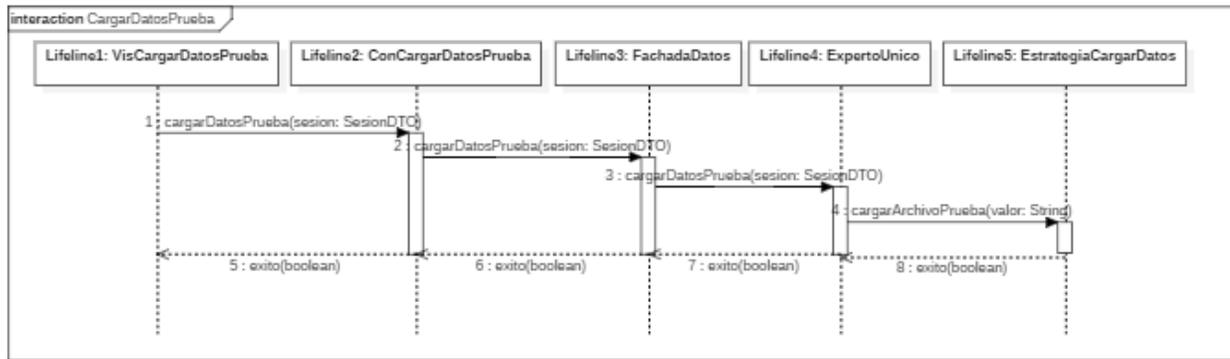


Figura 43. Diagrama de secuencia: Cargar datos de prueba. Autores de proyecto

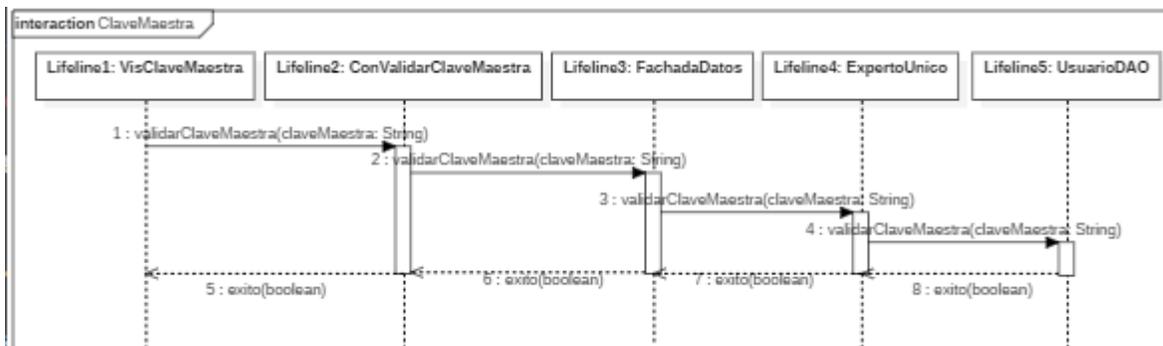


Figura 44. Diagrama de secuencia: Clave maestra. Autores de proyecto

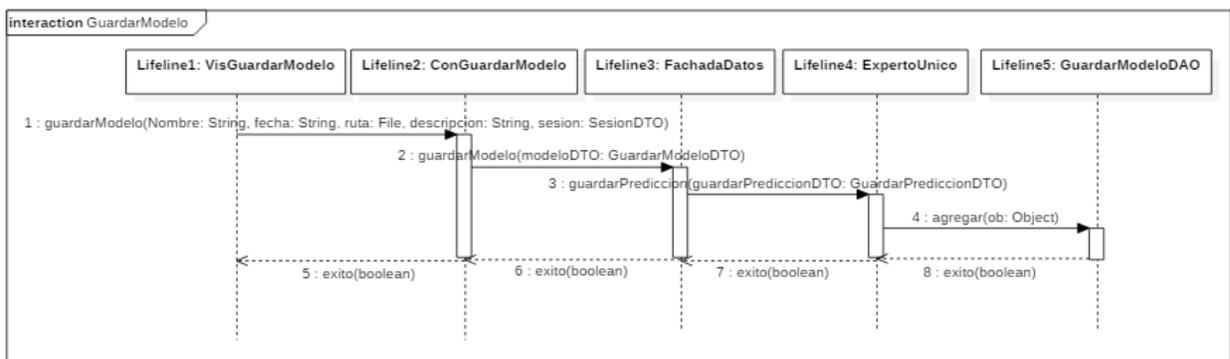


Figura 45. Diagrama de secuencia: Guardar modelo. Autores de proyecto

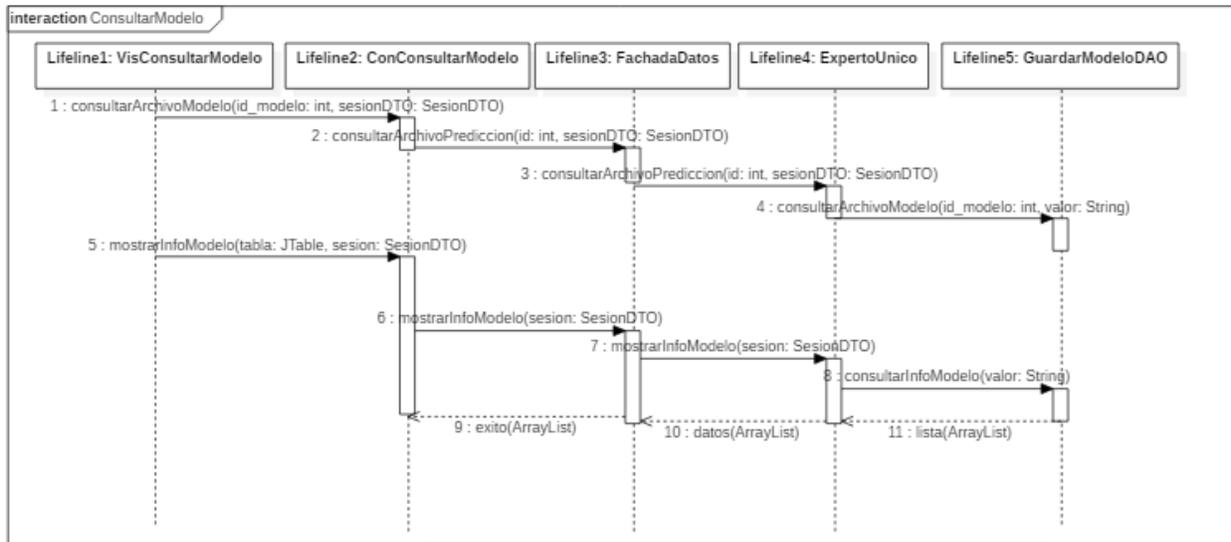


Figura 46. Diagrama de secuencia: Consultar modelo. Autores de proyecto

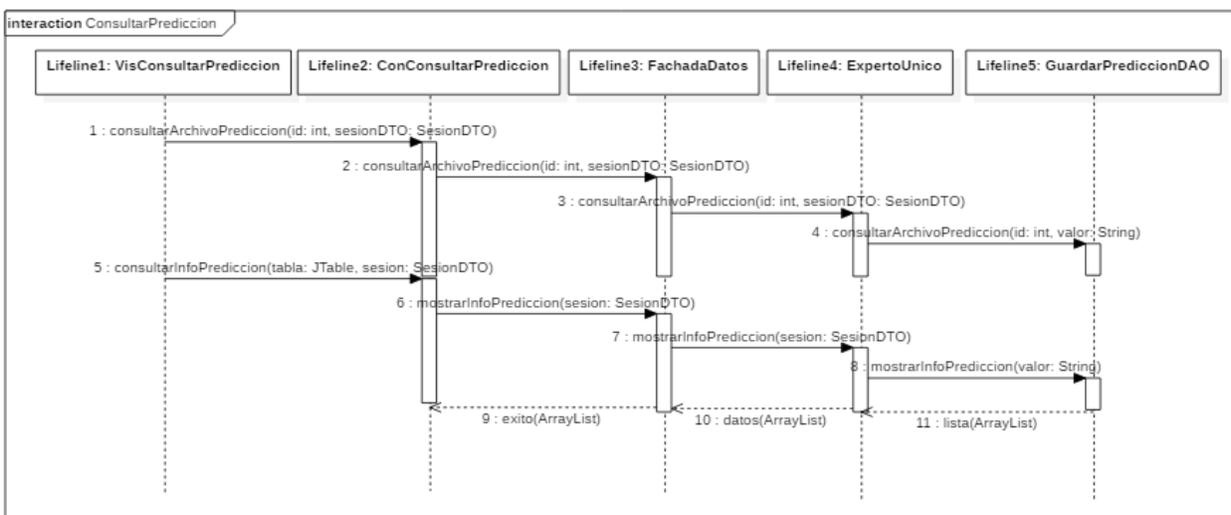


Figura 47. Diagrama de secuencia: Consultar predicción. Autores de proyecto

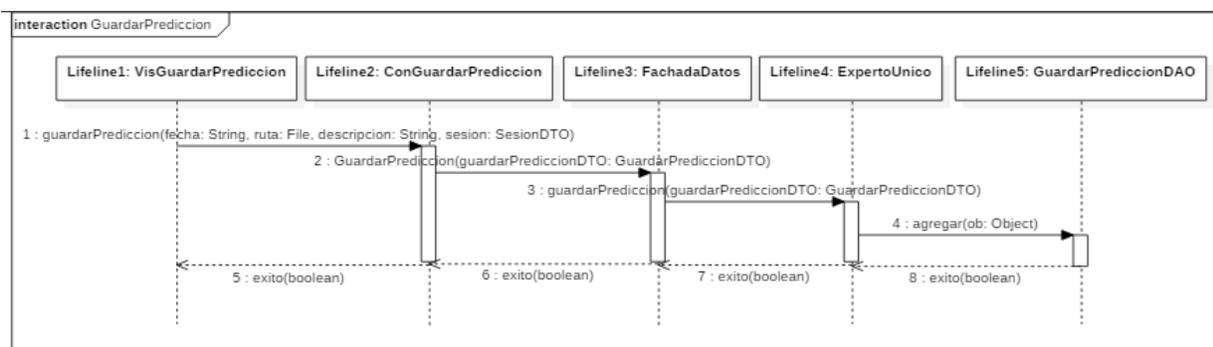


Figura 48. Diagrama de secuencia: Guardar predicción. Autores de proyecto

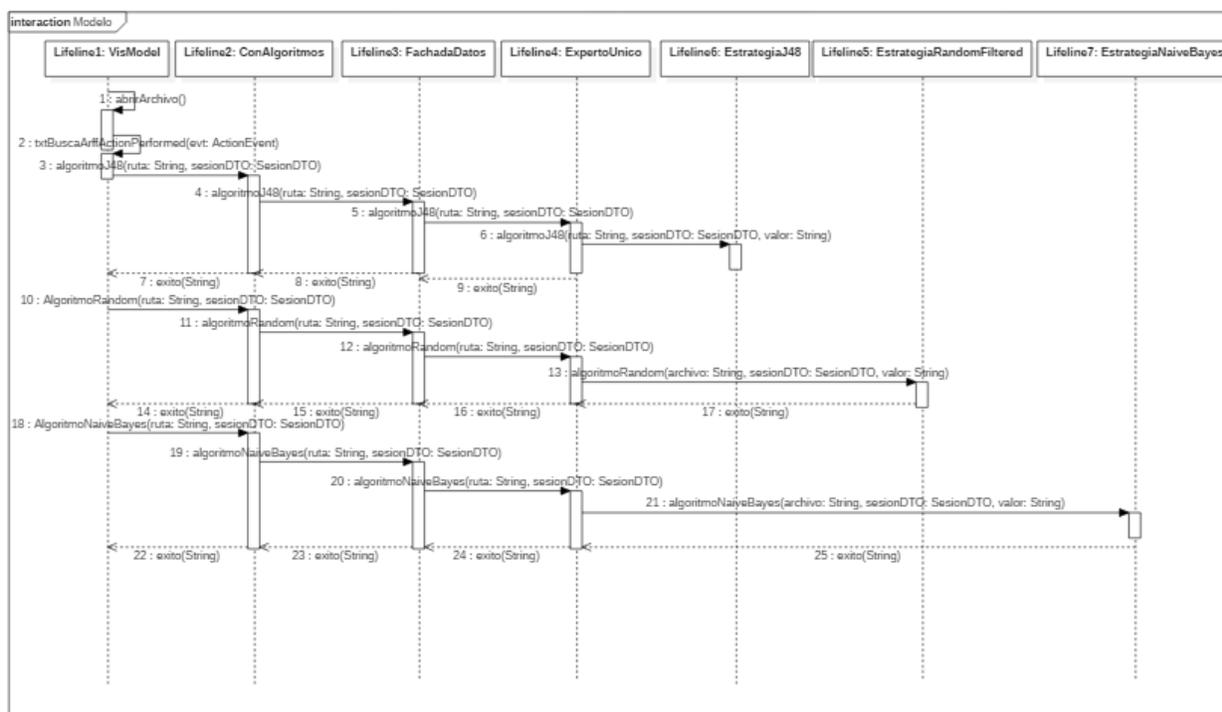


Figura 49. Diagrama de secuencia: Modelo. Autores de proyecto

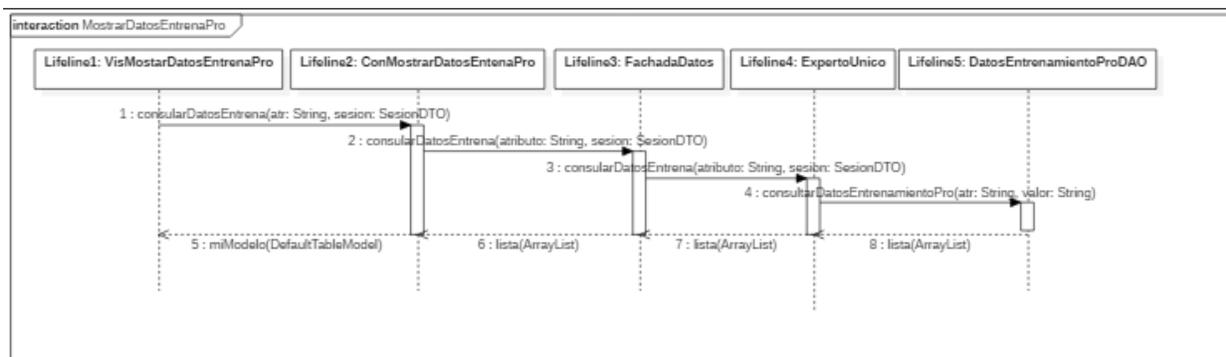


Figura 50. Diagrama de secuencia: Mostrar datos de entrenamiento. Autores de proyecto

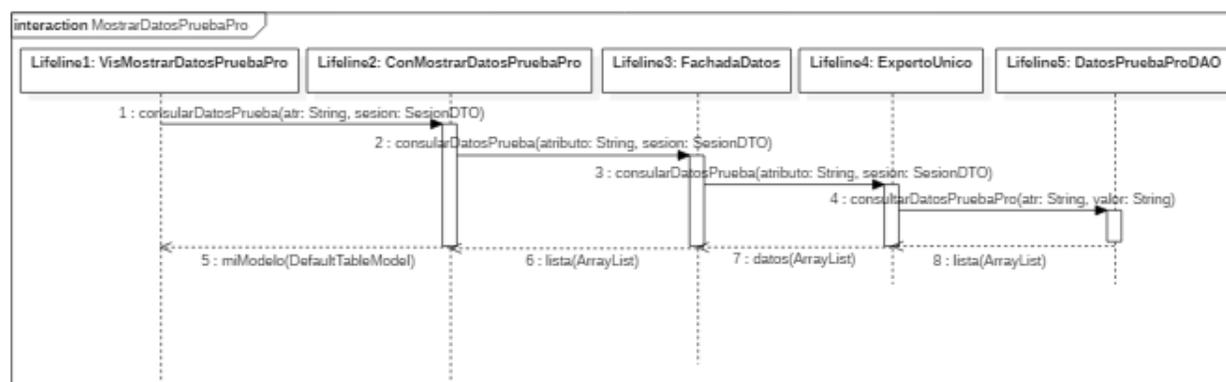


Figura 51. Diagrama de secuencia: Mostrar datos de prueba. Autores del proyecto

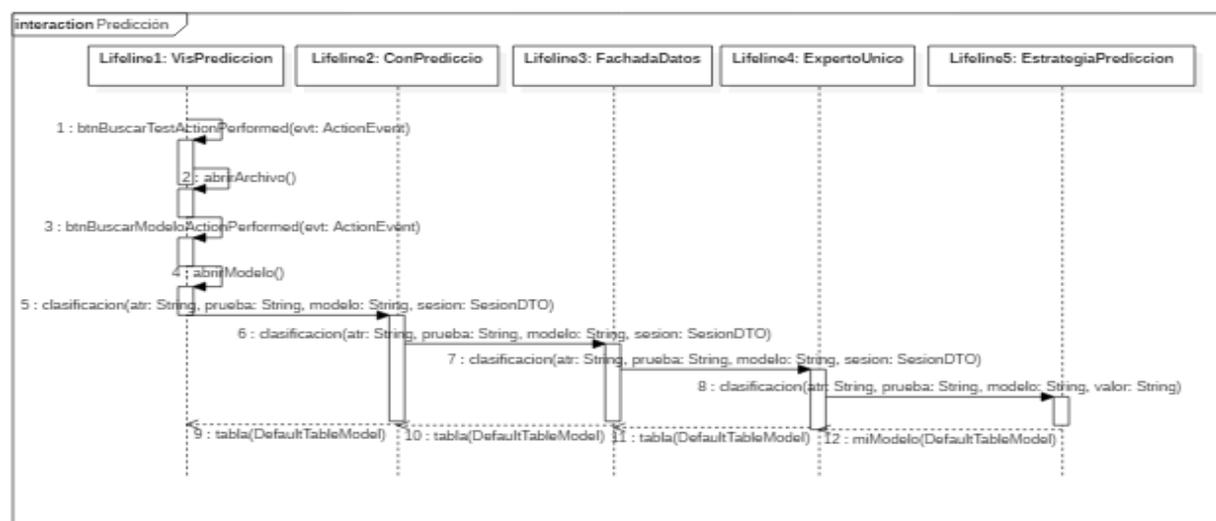


Figura 52. Diagrama de secuencia: Predicción. Autores de proyecto

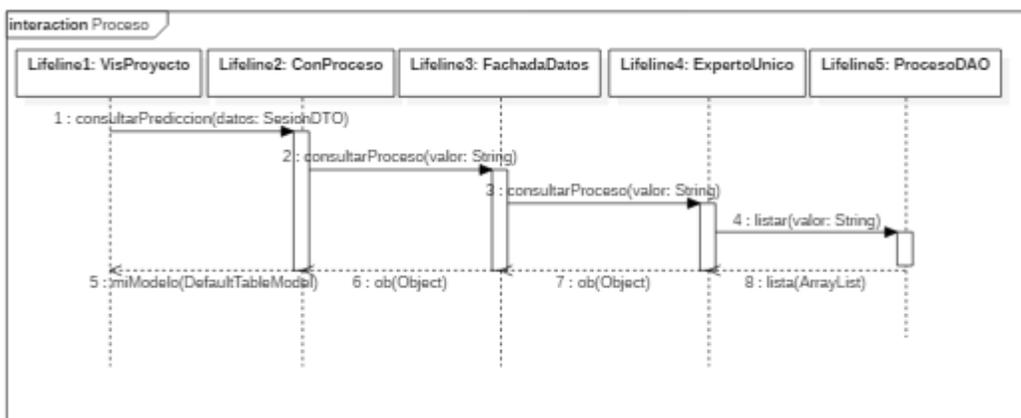


Figura 53. Diagrama de secuencia: Proceso. Autores de proyecto

4.3.2.3.6 Detalles del software.

En este punto se especifica la cantidad de líneas de código en cada paquete y cada clase. Finalmente se realiza el total de las líneas del software. Además de especificar las pruebas de integración más relevantes dentro del desarrollo del software.

Tabla 43
Pruebas de integración

N° de caso	Descripción del problema	Tratamiento del problema
1	Lentitud en el cargue de datos	Creación de un sql anidado con todas las discretizaciones del set de datos.
2	Valores de discretización faltantes en la clase promedio	Se creó un lista, colocando en esta los valores de discretización que le correspondían a la clase promedio
3	Al realizar el proceso del cargue de datos se detectó que cuando ocurría algún error se insertaban los datos en algunas tablas dejando otras vacías.	Para el tratamiento del caso anterior fueron las transacciones quienes evitaron este inconveniente, de modo que solo se insertaran los datos si se ejecutaban correctamente todas las consultas sql dando así, un <i>commit</i> , en caso contrario <i>rollback</i> .
4	Error al encontrar la ruta de la carpeta	Se agregó <i>user.dir</i> a la ruta de la carpeta.
5	Permitir que un proceso se ejecute varía veces del KDD.	Se crear una clase llamada <i>sesión</i> , el cual contiene la clase del usuario y clase del proceso, el cual dicho objeto se pasea por todas las clases y permite que un usuario

		solo pueda tener un ciclo de vida único en cada proceso.
6	No se especificaba a que estudiante pertenecía cada predicción	Por medio de un sql, se insertó y se llamó en un ArrayList dentro de un arreglo que generaba la tabla predicción.
7	Variación de información en el set de datos	Modificar el algoritmo en la estrategia que corresponda al set de datos.

Nota. Fuente: Autores del proyecto

Tabla 44

Líneas de código: paquetes Vistas

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
1	Vista	VisAgregarProceso	352
		VisArffEntrenamiento	280
		VisArffPrueba	268
		VisCargarDatosEntrenamiento	255
		VisCargarDatosPrueba	240
		VisClaveMaestra	157
		VisConsultarModelo	180
		VisConsultarPrediccion	184
		VisGuardarModelo	333
		VisGuardarPrediccion	251
		VisLogin	333
		VisModel	295
		VisMostrarDatosEntrenaPro	279
		VisMostrarDatosPruebaPro	278
		VisPrediccion	413
		VisPrincipal	664
		VisProceso	223
		VisUsuario	373
		TextPrompt	206
	Total de líneas de código		5561

Nota. Fuente: Autores del proyecto

Tabla 45
Líneas de código: paquete Controlador

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
2	Controlador	ConAgregarProceso	46
		ConAlgoritmos	90
		ConCargarDatosEntrena	45
		ConConsultarModelo	43
		ConConsultarPrediccion	147
		ConConvertidorEntrenamiento	108
		ConConvertidorPrueba	115
		ConGuardarModelo	70
		ConGuardarPrediccion	68
		ConInsertarUsuario	77
		ConLoginUsuario	60
		ConMostrarDatosEntenaPro	82
		ConMostrarDatosPruebaPro	78
		ConPrediccion	68
		ConProceso	146
		ConValidarClaveMaestra	39
Total de líneas de código			1282

Nota. Fuente: Autores del proyecto

Tabla 46
Líneas de código: paquete experto

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
3	Experto	expertoUnico	528
Total de líneas de código			528

Nota. Fuente: Autores del proyecto

Tabla 47
Líneas de código: paquete fachada

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
4	Fachada	FachadaDatos	449
Total de líneas de código			449

Nota. Fuente: Autores del proyecto

Tabla 48
Líneas de código: paquete modelo

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
5	Modelo.dao	DatosEntrenamientoDAO	360
		DatosEntrenamientoProDAO	238
		DatosPruebaDAO	360
		DatosPruebaProDAO	281
		GuardarModeloDAO	249
		GuardarPrediccionDAO	213
		Hash	67
		ProcesoDAO	189
		UsuarioDAO	245
		6	Modelo.dto
DatosPruebaDTO	250		
GuardarModeloDTO	99		
GuardarPrediccionDTO	81		
MostrarDatosEntrenamientoProDTO	39		
MostrarDatosPruebaProDTO	38		
ProcesoDTO	122		
SesionDTO	271		
UsuarioDTO	140		
7	Modelo.estrategia		
		EstrategiaConvertidor	220
		EstrategiaJ48	105
		EstrategiaNaiveBayes	96
		EstrategiaPrediccion	188
		EstrategiaRandomFiltered	98
8	Modelo.interface2	ICRUD	17
9	Modelo.conexion	Conectar	96
Total de líneas de código			4662

Nota. Fuente: Autores del proyecto

Tabla 49
Líneas de código: paquete util

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
10	Útil	Util	61
Total de líneas de código			61

Nota. Fuente: Autores del proyecto

Tabla 50
Líneas de código: paquete principal

Consecutivo	Nombre de paquetes	Nombre de clase	Líneas de código
11	Principal	Principal	26
Total de líneas de código			26

Nota. Fuente: Autores del proyecto

Tabla 51
Total de líneas de código software

Consecutivo	Nombre del paquete	Líneas de código
1	Vista	5561
2	Controlador	1282
3	Experto	528
4	Fachada	449
5,6,7,8,9	Modelo	446
10	Útil	61
11	Principal	66
Total de líneas de código		8393

Nota. Fuente: Autores del proyecto

El software cuenta con un total de 12569 líneas de código, el cual cada una de ellas es de vital importancia para el funcionamiento exitoso del software.

4.3.2.3.7 Manuales del software. Manual del sistema: en este manual se detalla el código fuente del software, con el fin de que si se realizan nuevas modificaciones sobre este, el desarrollaron encargado pueda conocer la información necesaria sobre cada método, variable o cualquier elemento dentro del sistema.

- ❖ Manual de instalación: este manual se encarga de guiar al usuario en la instalación del sistema, que inicia desde la creación de la base de datos hasta la instalación del ejecutable del sistema (ver Apéndice J).

- ❖ Manual del usuario: el manual del usuario especifica las interfaces con las que interactúa el usuario, detallando cada acción del sistema, con el fin de guiar al usuario en la trazabilidad del sistema, que inicia con el login hasta generar la predicción, siendo este el objetivo central del software (ver Apéndice K).

4.3.2.4 Transición Fase Iteración. En esta fase se establece la prueba beta para validar que se cumplan las expectativas del usuario; los resultados en dicha prueba están representados en el manual de usuario (ver Apéndice K), donde se detalla el correcto funcionamiento del sistema, la ruta que se estipula dentro del manual de usuario es la que se mostró a las partes interesadas, las cuales quedaron satisfechas con el resultado final del proyecto.

Finalmente, se aprecia en el desarrollo de las fases de la metodología open up la agilidad de dicha metodología, permitiendo crear en el tiempo adecuado los diferentes artefactos vistos anteriormente, también es importante resaltar que con el uso de esta metodología se ha podido llegar al objetivo que se tenía con ella, el cual era la creación de un software que permitiera detectar índices de deserción en el programa de ingeniería de sistemas.

4.4. Análisis e interpretación de los resultados obtenidos

El desarrollo de las fases señaladas a cumplir dentro de la metodología cualitativa con enfoque descriptivo que se planteó para el proyecto de investigación no se tuvo ningún inconveniente y gracias al detalle de cada metodología en el desarrollo completo del proyecto se puede observar lo bien estructurado que se han definido cada una de ellas, permitiendo a través de esta el correcto cumplimiento de los objetivos generales de la investigación y el total desarrollo del objetivo general de la misma.

Capítulo 5. Conclusiones

En la selección de las variables que se usaron dentro de la investigación fue de gran ayuda la revisión literaria, pues con esta, se ha permitido tener una visión global de cómo se abarca el tema de la deserción estudiantil a nivel regional, nacional e internacional.

En la aplicación del proceso de extracción del conocimiento (KDD) se han podido establecer de forma detallada todos los elementos necesarios para predicción tomando como inicio lo relacionado a los datos y finalizando con la predicción de la clase promedio a través de un modelo generado en el proceso.

La construcción de una herramienta que permite la automatización de todo el proceso KDD es de gran ayuda, ya que permite agilizar tareas de minería de datos. En este sentido, la elaboración del software cuenta con todos los elementos necesarios, en donde se aplica la técnica de clasificación, para la predicción del rendimiento académico de los estudiantes.

Finalmente, los resultados obtenidos con la herramienta son los que se esperaron en el inicio de la investigación, pues se ha podido desarrollar un software que permite la predicción del promedio que tendrá cada estudiante, teniendo en cuenta que el rendimiento académico es considerado como un tipo de abandono, además de ser uno de los factores más influye en la deserción estudiantil.

Capítulo 6. Recomendaciones

1. Se recomienda comprobar que java este instalado en el equipo donde se realizara la instalación del software, ya que es necesario tener instalado el archivo *jre* para el correcto funcionamiento del software.
2. También se recomienda que al momento de solicitar los datos se respete el formato del archivo y la estructura de las columnas.
3. Se recomienda reinstalar el gestor de base de datos: XAMPP, pues se necesita de recursos con lo que cuenta la última actualización de este gestor.
4. En relación a los datos suministrados por parte del módulo con el que actualmente división de sistemas le suministro al plan de estudios de ingeniería de sistemas, se recomienda que este archivo sea dividido en dos partes, una que corresponda a los datos de los estudiantes matriculados en primer semestre dentro de la carrera y la otra parte formen parte del histórico de datos (entrenamiento).
5. Se recomienda que el software se use al inicio de cada semestre con los estudiantes que inician su periodo académico dentro de la carrera de ingeniería de sistemas.

6. Con el tratamiento de los datos, suministrados por parte del sistema de admisiones, registro y control, se pudieron encontrar las siguientes inconsistencias:

6.1 Códigos estudiantiles repetidos: se recomienda hacer una validación al momento de generar los códigos de cada estudiante, la cual evite que se registre en el sistema un código que ya este asignado a otro estudiante.

6.2 El estado alumno según los datos suministrados tiene las siguientes categorías: activo, inactivo y excluido, pero, en el análisis de esta información se ha detectado que los estudiantes que pertenecen a la categoría de inactivos son de diferente tipo, pues se detectaron como inactivos códigos de estudiantes que han cambiado de pensum, que no han pagado matrícula académica y estudiantes que han culminado la carrera. Para la solución a este inconveniente es necesario una nueva discretización de los datos, donde, se amplíe la categoría de los mismos, o crear una nueva columna que contenga esa nueva discretización, con el fin de que esta sea la nueva clase de predicción.

6.3 Algunos estudiantes tenían el campo de valoración icfes vacío, esto es una validación que se debería dar en la inserción de los datos, donde no se permita guardar este campo vacío, pues la admisión de un estudiante al programa de ingeniería de sistemas es por medio de este puntaje, a menos que sean estudiantes de intercambio.

6.4 La ocupación es una variable importante dentro de la minería en el tema de deserción estudiantil, pero no se pudo hacer uso de esta, debido a la variabilidad de los datos, pues son insertados a la base de datos por medio de un campo de texto. Lo que se recomienda es que el campo de ocupación se estandarice según la estructura de la clasificación

internacional uniforme de ocupaciones, con el fin de que no genere ruido la variabilidad de los datos, y se tengan que eliminar al momento de revisar la calidad de los datos.

6.5 Los campos que indica si el estudiante pertenece a una etnia, si es desplazado y conflicto se recomienda que se almacenen por medio de selección, pues los datos que han sido suministrados en esta columna no van acorde a este tipo.

Referencias

- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Advanced Materials Research*, (65), 9–36. <https://doi.org/10.4028/www.scientific.net/AMR.699.273>
- Crotti, E., Armas, G. De, Ketele, J. De, Gajardo, M., García-huidobro, J. E., & María, J. (2004). *Políticas educativas y equidad*.
- Chen, R. (2012). Institutional Characteristics and College Student Dropout Risks: A Multilevel Event History Analysis. *Research in Higher Education*, 53(5), 487–505. <https://doi.org/10.1007/s11162-011-9241-4>
- Gartner, M. L., & Gallego, C. (2015). La deserción estudiantil en la Universidad de Caldas: sus características, factores determinantes y el impacto de las estrategias institucionales de prevención. V *CLABES Quinta Conferencia Latinoamericana Sobre El Abandono En La Educación Superior*. Retrieved from http://www.alfaguia.org/www-alfa/images/ponencias/clabesv/L1_Ponencias/5_CLABES_paper_55.pdf
- Gomez, D., & Velasquez, E. (2017). IDENTIFICACIÓN DE LAS VARIABLES QUE MOTIVAN LA DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE ADMINISTRACIÓN DE EMPRESAS DE LA UFPSO. *Tesis*, 1–49.
- Henríquez, N., & Escobar, D. (2016). Construcción de un modelo de alerta temprana para la detección de estudiantes en riesgo de deserción de la universidad metropolitana de ciencias de la educación. *Revista Mexicana de Investigación Educativa*, 21(71), 1221–1248.
- Reyes, J; Escobar, C; Duarte, J & Ramírez, P (2007). Una aplicación del modelo de regresión logística en la predicción del rendimiento estudiantil. *Estudios Pedagógicos XXXIII,Nº2*, pp: 101-120.
- Saldaña Villa, M., & Barriga, O. A. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile TT - An Adaptation of Tinto's Attrition Model to the Universidad Católica de la Santísima Concepción, Chile. *Revista de Ciencias Sociales*, 16(4), 616–628. Retrieved from http://www.scielo.org.ve/pdf/rcs/v16n4/art05.pdf%5Cnhttp://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S1315-95182010000400005<=pt
- Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, 33(1), 107–124. <https://doi.org/10.1111/exsy.12135>
- Mayra, A., & Mauricio, D. (2018). Factors to predict dropout at the universities: A case of study in Ecuador. *IEEE Global Engineering Education Conference, EDUCON, 2018–April*, 1238–1242. <https://doi.org/10.1109/EDUCON.2018.8363371>
- MEN. (2014). *SPADIES (Sistema Para la Prevención de la Educación de la Educación Superior)*. Obtenido de <https://www.mineduacion.gov.co/sistemasdeinformacion/1735/w3-article-254648.html>
- Mineduacion. (2017). Reporte Sobre Deserción y Graduación en Educación Superior, 57(57), 1–4. Retrieved from <https://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles->

- 357549_recurso_5.pdf
- Mishra, T., Kumar, D., & Gupta, S. (2014). Mining students' data for prediction performance. *International Conference on Advanced Computing and Communication Technologies, ACCT*, 255–262. <https://doi.org/10.1109/ACCT.2014.105>
- Gartner Isaza, M. L., & Gallego Giraldo, C. (2015). La deserción estudiantil en la Universidad de Caldas: sus características, factores determinantes y el impacto de las estrategias institucionales de prevención. *Quinta Conferencia latinoamericana sobre el abandono en la educación superior*. Talca.
- Guzmán Ruiz, C., Durán Muriel, D., & Franco Gallego, J. (2009). *Deserción estudiantil en la educación superior de Colombia*. Bogota.
- Reyes, L., Carranza, E. C., & Castro, D. P. (2012). Causas psicosociales de la deserción universitaria. *Revista LOGOS CIENCIA & TECNOLOGÍA ISSN*, 4(1), 2145–2549. Retrieved from <http://www.redalyc.org/pdf/5177/517751763013.pdf>
- Saldaña Villa, M., & Barriga, O. A. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile TT. *Revista de Ciencias Sociales*, 16(4), 616–628. Retrieved from http://www.scielo.org.ve/pdf/rcs/v16n4/art05.pdf%5Cnhttp://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S1315-95182010000400005(=pt
- Suárez, N., Luz, Y., & Díaz, B. (2015). Estrés académico, deserción y estrategias de retención de estudiantes en la educación superior. *Rev. Salud Pública*, 17(2), 300–313. <https://doi.org/10.15446/rsap.v17n2.52891>
- Universidad de los Andes. (2014). Determinantes de la deserción. <https://doi.org/10.4018/978-1-4666-4979-8>
- Vélez, Amparo; López Jiménez, D. F. (2009). Estrategias para vencer la deserción universitaria. *Educación y Educadores*, 7(0123-1294), 177–203. Retrieved from <http://educacionyeducadores.unisabana.edu.co/index.php/eye/article/view/555>
- Yolima, I., Rojas, V., Pablo, J., Torres, S., & Temática, L. (2016). El Abandono Estudiantil En La Universidad Simón Bolívar: Características Y Factores Determinantes En La Frontera Colombia Y Venezuela. *VI CLABES. Sexta Conferencia Sobre El Abandono En La Educación Superior*.
- Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). EDM-62: Application of Data Mining Methods in an Online Education Program. *European Journal of Open, Distance and E-Learning*, 17(1), 118–133. <https://doi.org/10.2478/eurodl-2014-0008>
- Zarate, R., & Manilla, E. (2014). La deserción estudiantil UIS, una mirada desde la responsabilidad social universitaria.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004). Deserción Estudiantil Universitaria, (60), 39–65.
- JAVA. (2018). *Oracle Corporation*. Obtenido de https://www.java.com/es/download/faq/whatis_java.xml
- MINEDUCACION. (2018). *MINISTERIO DE EDUCACIÓN*. Obtenido de <https://www.mineducacion.gov.co/1621/article-82745.html>
- MINEDUCACION. (2018). *MINISTERIO DE EDUCACIÓN*. Obtenido de <https://www.mineducacion.gov.co/1621/article-80317.html>
- Gallego, L., Castrillón, W., & Marín, J. (2017). FACTORES QUE INTERVIENEN EN LA INTENCIONALIDAD DE DESERCIÓN ESTUDIANTIL EN EL PROGRAMA DE

- ZOOTECNIA DE LA UNIVERSIDAD DE ANTIOQUIA, SEDE MEDELLÍN. *Séptima Conferencia Latinoamericana Sobre El Abandono En La Educación Superior*.
- Fayyad, U., Piatestky-Shapiro, G. & Smyth, P. (1996). The kdd Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the acm*, 39(11), 27-34.
- Altamiranda, J., Aguilar, J., & Hernandez, L. (2015). Sistema de reconocimiento de patrones de sustancias químicas cerebrales basado en minería de datos. *Computacion y Sistemas*, 19(1), 89–107. <https://doi.org/10.13053/CyS-19-1-1409>
- Anaya, A. T. (2007). Introducción Software. Retrieved from http://www.uaeh.edu.mx/docencia/P_Presentaciones/prepa4/informatica/Software_1.pdf
- ANGEL, A. J. J., & GRANADOS, R. D. (2013). ANALISIS DE LA POLITICA ANTI-DESERCIÓN ESTUDIANTIL EN TECNAR. *Biblioteca Universitaria*.
- Angulo, F., & Sergio, E. (2012). Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios, 95. <https://doi.org/10.19052/ap.4075>
- BECERRA, A. L. B., & IRAL, P. C. (2008). EVALUACION DE LAS ESTRATEGIAS PARA DISMINUIR LA DESERCIÓN Y LA REPITENCIA EN EL PROGRAMA DE LICENCIATURA EN PEDAGOGIA INFANTIL. *Biblioteca Universitaria*.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Advanced Materials Research*, (65), 9–36. <https://doi.org/10.4028/www.scientific.net/AMR.699.273>
- Castorena Peña, J. A., Silva Ávila, A. E., Domínguez Lugo, A. J., & Rodríguez Montelongo, D. L. (2018). El uso de herramientas tecnológicas de minería de datos en el análisis de datos climatológicos / The use of technological tools for data mining in the analysis of climatological data. *RECI Revista Iberoamericana de Las Ciencias Computacionales e Informática*, 7(13), 1. <https://doi.org/10.23913/reci.v7i13.75>
- Cristina, I., Gutiérrez, M., María, P., Hurtado, A., Cardona, S. G., Iván, F., ... Zea, E. T. (2010). Analisis de la Deserción Estudiantil en los programas de pregrado de la universidad EAFIT. *Cuadernos de Investigación*, 118. Retrieved from http://www.alfaguia.org/alfaguia/files/1320244234_30.pdf
- Ferreya, M., Avitabile, C., Botero, J., Haimovich, F., & Urzúa, S. (2017). *Momento decisivo (La educación superior en América Latina y el Caribe)*.
- Franzante, B. A. (2013). Permanencia estudiantil en el contexto universitario actual. Reto pedagógico para la igualdad de oportunidades educativas. *Congreso Universidad*, III(2), 1–9.
- Gartner, M. L., & Gallego, C. (2015). La deserción estudiantil en la Universidad de Caldas: sus características, factores determinantes y el impacto de las estrategias institucionales de prevención. *V CLABES Quinta Conferencia Latinoamericana Sobre El Abandono En La Educación Superior*. Retrieved from http://www.alfaguia.org/www-alfa/images/ponencias/clabesv/L1_Ponencias/5_CLABES_paper_55.pdf
- Gomez, D., & Velasquez, E. (2017). IDENTIFICACIÓN DE LAS VARIABLES QUE MOTIVAN LA DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE ADMINISTRACIÓN DE EMPRESAS DE LA UFPSO. *Tesis*, 1–49.
- Guzmán, C., Muriel, D., Nacional, Franco, J., Castaño, E., Gómez, K., & Vásquez, J. (2009). *Desercion estudiantil en la Educación Superior Colombiana*.
- Henríquez, N., & Escobar, D. (2016). Construcción de un modelo de alerta temprana para la detección de estudiantes en riesgo de deserción de la universidad metropolitana de ciencias de la educación. *Revista Mexicana de Investigacion Educativa*, 21(71), 1221–1248.

- Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación superior. *Retención y Movilidad Estudiantil*. <https://doi.org/10.31619/caledu.n17.409>
- Kerbo, F. H. R., Hill, M., Sistema, E. L., Mundial, D. E. E., & Hill, M. (1991). ESTRUCTURA SOCIAL . ESTRUCTURA SOCIAL . Fuente : Harold R . Kerbo , Estratificación social y, 1–15.
- Lopera, C. (2008). Determinantes de la deserción universitaria en la Facultad de Economía Universidad del Rosario. *Borradores de Investigación de La Universidad Del Rosario*, (95), 26.
- Manament Solutions. (2018). Machine Learning una pieza clave en la transformación de los modelos de negocio. Retrieved from <https://www.managementsolutions.com/sites/default/files/publicaciones/esp/machine-learning.pdf>
- Marcano Aular, Y. Josefina, & Talavera Pereira, R. (2007). Minería de datos como soporte a la toma de decisiones empresariales. *Opción*, 23(52), 104–118.
- Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, 33(1), 107–124. <https://doi.org/10.1111/exsy.12135>
- Mayra, A., & Mauricio, D. (2018). Factors to predict dropout at the universities: A case of study in Ecuador. *IEEE Global Engineering Education Conference, EDUCON, 2018–April*, 1238–1242. <https://doi.org/10.1109/EDUCON.2018.8363371>
- MELLENDEZ, R. (2008). ESTUDIO SOBRE DESERCIÓN Y PERMANENCIA ACADÉMICA EN LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE LA GUAJIRA DESDE EL II PA 2005 HASTA EL II PA 2007. *Repositorio Institucional*, 282.
- MEN. (2012). MANUAL DE CONCEPTOS SISTEMAS DE INFORMACIÓN SPADIES INVESTIGACIÓN. *SPADIES*.
- MEN. (2017). Reporte Sobre Deserción y Graduación en Educación Superior, 57(57), 1–4. Retrieved from https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articulos-357549_recurso_5.pdf
- Ministerio de Educación Nacional. (2014). Acuerdo Nacional para la disminución de la deserción en la educación superior, 57(1), 17. <https://doi.org/10.1016/j.jvolgeores.2017.06.019>
- Mishra, T., Kumar, D., & Gupta, S. (2014). Mining students' data for prediction performance. *International Conference on Advanced Computing and Communication Technologies, ACCT*, 255–262. <https://doi.org/10.1109/ACCT.2014.105>
- Moreno, D. M. (2013). La Deserción Escolar: Un problema de Carácter Social. *Revista In Vestigium Ire*, 6, 115–124. Retrieved from <http://revistas.ustatunja.edu.co/index.php/ivestigium/article/download/795/716>
- Quintela-Dávila, G. (2013). Deserción universitaria, una aproximación sociológica al proceso de toma de decisiones de los estudiantes. *Sociedad Hoy*, 24, 83–106.
- Reyes, L., Carranza, E. C., & Castro, D. P. (2012). Causas psicosociales de la deserción universitaria. *Revista LOGOS CIENCIA & TECNOLOGÍA ISSN*, 4(1), 2145–2549. Retrieved from <http://www.redalyc.org/pdf/5177/517751763013.pdf>
- Roger S. Pressman. (2010). *Ingeniería de software - Un Enfoque práctico*. *Danielr.Obolog.Es*. <https://doi.org/http://zeus.inf.ucv.cl/~bcrawford/Modelado%20UML/Ingenieria%20del%20Software%207ma.%20Ed.%20-%20Ian%20Sommerville.pdf>
- Rosado, A., & Verjel, A. (2017). APLICACIÓN DE LA MINERÍA DE DATOS EN LA

- EDUCACION EN LINEA. *Revista Colombiana de Tecnologías de Avanzada*.
- Rueda, R. Z., & Socha, C. M. (2009). Estudio sobre las motivaciones de deserción estudiantil en la Universidad Industrial de Santander.
- Saldaña Villa, M., & Barriga, O. A. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile TT. *Revista de Ciencias Sociales*, 16(4), 616–628. Retrieved from http://www.scielo.org.ve/pdf/rcs/v16n4/art05.pdf%5Cnhttp://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S1315-95182010000400005(=pt
- Suárez, N., Luz, Y., & Díaz, B. (2015). Estrés académico, deserción y estrategias de retención de estudiantes en la educación superior. *Rev. Salud Pública*, 17(2), 300–313. <https://doi.org/10.15446/rsap.v17n2.52891>
- Timarán Pereira, R., & Yépez Chamorro, M. C. (2012). Applied data mining patterns to discovery survival in women with invasive cervical cancer. *Universidad y Salud*, 14(2), 117–129. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0124-71072012000200002&lng=es&nrm=iso&tlng=es
- Universidad de los Andes. (2014). Determinantes de la deserción. <https://doi.org/10.4018/978-1-4666-4979-8>
- Vásquez, R. J. O. (2015). Metodología para la detección temprana de la deserción de los estudiantes del pregrado sistemas de información en la asignatura lógica y programación del Instituto Tecnológico Metropolitano., 107. <https://doi.org/http://dx.doi.org/10.1016/j.fuel.2004.01.013>
- Vélez, Amparo; López Jiménez, D. F. (2009). Estrategias para vencer la deserción universitaria. *Educación y Educadores*, 7(0123-1294), 177–203. Retrieved from <http://educacionyeducadores.unisabana.edu.co/index.php/eye/article/view/555>
- Yolima, I., Rojas, V., Pablo, J., Torres, S., & Temática, L. (2016). El Abandono Estudiantil En La Universidad Simón Bolívar: Características Y Factores Determinantes En La Frontera Colombia Y Venezuela. *VI CLABES. Sexta Conferencia Sobre El Abandono En La Educación Superior*.
- Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). EDM-62: Application of Data Mining Methods in an Online Education Program. *European Journal of Open, Distance and E-Learning*, 17(1), 118–133. <https://doi.org/10.2478/eurodl-2014-0008>
- Zarate, R., & Manilla, E. (2014). La deserción estudiantil UIS, una mirada desde la responsabilidad social universitaria.
- Daniel, A. (2001). ARQUITECTURAS Y TECNOLOGÍAS PARA EL DESARROLLO DE APLICACIONES WEB. *Arquitecturas y Tecnologías para el Desarrollo de Aplicaciones Web*.
- Eclipse. (agosto de 2007). *Introduction to OpenUP (Open Unified Process)*. Obtenido de <https://www.eclipse.org/epf/general/OpenUP.pdf>
- García Gutiérrez, J. (2016). *Comenzando con Weka: Filtrado y selección de subconjuntos de atributos basados en su relevancia descriptiva para la clase*.
- García Morate, D. (s.f.). *Doctoral Programme Education in the Knowledge Society*. Obtenido de <https://knowledgesociety.usal.es/sites/default/files/MANUAL%20WEKA.pdf>
- Gómez Lopez, L. F. (2005). Unión de universidades de América Latina y el Caribe. *Universidades*, 29, 39-45.

- Guzman, G., & Tovar, C. (marzo de 2013). IMPLEMENTACION DEL PATRON DE MVC PARA EL PROCESO DE SELECCIÓN DE PERSONAL. *Repositorio de la Universidad Libre*.
- Muñoz Reja, I. C., Gómez Cejudo, A. I., Gualo García, J., Rivas García, B., & Piattini Velthuis, M. (2019). *Calidad de datos*. Bogotá: Ra-ma.
- Rodríguez Rodríguez, J. E. (2010). *Fundamento de minería de datos*. Bogotá: Universidad distrital francisco José de Caldas, ASEUC.
- Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, S. J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en base de datos. En S. R. Timarán Pereira, I. Hernández Arteaga, S. J. Caicedo Zambrano, A. Hidalgo Troya, & J. C. Alvarado Pérez, *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (págs. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia.

Apéndices

Apéndice A. Artículo

Este Apéndice, contiene un artículo relacionado con una revisión acerca de las herramientas, técnicas, algoritmos y atributos que son más usados en el área de deserción estudiantil a nivel nacional e internacional (Review of techniques, tools, algorithms and attributes for data mining used in student desertion).

Apéndice B. Conjunto de datos

Corresponde a la base de datos suministrada por parte de la oficina de admisiones, registro y control (Datos_sistemas).

Apéndice C. Experimento 1

- El método de evaluación: *ClassifierSubsetEval*
- El método de búsqueda: *GreedyStepWise*

J48	Naïve Bayes	RandomizableFilteredClassifier
=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===	=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===	=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===
number of folds (%) attribute	number of folds (%) attribute	number of folds (%) attribute
5(100 %) 1 ESTADO_ALUMNO	1(20 %) 1 ESTADO_ALUMNO	5(100 %) 1 ESTADO_ALUMNO
5(100 %) 2 SANCION	5(100 %) 2 SANCION	5(100 %) 2 SANCION
4(80 %) 3 ANO_INGRESO	0(0 %) 3 ANO_INGRESO	5(100 %) 3 ANO_INGRESO
5(100 %) 4 SEMESTRE_ACTUAL	5(100 %) 4 SEMESTRE_ACTUAL	5(100 %) 4 SEMESTRE_ACTUAL
5(100 %) 5 VALORACION_ICFES	3(60 %) 5 VALORACION_ICFES	5(100 %) 5 VALORACION_ICFES
5(100 %) 6 EDAD	4(80 %) 6 EDAD	5(100 %) 6 EDAD
4(80 %) 7 SEXO	4(80 %) 7 SEXO	5(100 %) 7 SEXO
5(100 %) 8 ESTRATO_SOCIAL	0(0 %) 8 ESTRATO_SOCIAL	5(100 %) 8 ESTRATO_SOCIAL
4(80 %) 9 LUGAR_PROCEDENCIA	2(40 %) 9	5(100 %) 9
5(100 %) 10 DPTO_PROCEDENCIA	LUGAR_PROCEDENCIA	LUGAR_PROCEDENCIA
	0(0 %) DPTO_PROCEDENCIA	5(100 %) 10 DPTO_PROCEDENCIA

4(80 %) 11	0(0 %) 11	5(100 %) 11
NUMERO_HERMANOS_ESTUDIAN	NUMERO_HERMANOS_ESTUDIAN	NUMERO_HERMANOS_ESTUDIAN
0(0 %) 12 ESTADO_CIVIL	3(60 %) 12 ESTADO_CIVIL	2(40 %) 12 ESTADO_CIVIL

Apéndice D. Experimento 2

- El método de evaluación: *CfsSubsetEval*
- El método de búsqueda: *BestFirts*

=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===

number of folds (%) attribute

5(100 %)	1	ESTADO_ALUMNO
5(100 %)	2	SANCION
0(0 %)	3	ANO_INGRESO
5(100 %)	4	SEMESTRE_ACTUAL
0(0 %)	5	VALORACION_ICFES
0(0 %)	6	EDAD
2(40 %)	7	SEXO
0(0 %)	8	ESTRATO_SOCIAL
0(0 %)	9	LUGAR_PROCEDENCIA
2(40 %)	10	DPTO_PROCEDENCIA
0(0 %)	11	NUMERO_HERMANOS_ESTUDIAN
0(0 %)	12	ESTADO_CIVIL

Apéndice E. Experimento 3

- El método de evaluación: *OneRAttributeEval*
- El método de búsqueda: *Ranker*

=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===

average merit	average rank	attribute
0.516 +- 0.005	1 +- 0	2 SANCION
0.409 +- 0.008	2 +- 0	4 SEMESTRE_ACTUAL
0.329 +- 0.011	3 +- 0	1 ESTADO_ALUMNO
0.167 +- 0.006	4 +- 0	6 EDAD
0.12 +- 0.007	5.4 +- 0.49	3 ANO_INGRESO
0.116 +- 0.006	5.6 +- 0.49	5 VALORACION_ICFES
0.06 +- 0.005	7 +- 0	11 NUMERO_HERMANOS_ESTUDIAN
0.02 +- 0.002	8.8 +- 0.75	12 ESTADO_CIVIL
0.02 +- 0.003	9 +- 0.89	8 ESTRATO_SOCIAL
0.017 +- 0.004	9.4 +- 1.02	10 DPTO_PROCEDENCIA
0.012 +- 0.002	10.8 +- 0.4	7 SEXO
0.005 +- 0.002	12 +- 0	9 LUGAR_PROCEDENCIA

Apéndice F. Experimento 4

- El método de evaluación: *CorrelationAttributeEval*
- El método de búsqueda: *Ranker*

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 13 ESTADO_PROMEDIO):

Correlation Ranking Filter

Ranked attributes:

0.4063 2 SANCION
 0.2054 1 ESTADO_ALUMNO
 0.1803 4 SEMESTRE_ACTUAL
 0.0873 3 ANO_INGRESO
 0.0784 5 VALORACION_ICFES
 0.0761 11 NUMERO_HERMANOS_ESTUDIAN
 0.0501 6 EDAD
 0.0498 10 DPTO_PROCEDENCIA
 0.048 7 SEXO
 0.043 8 ESTRATO_SOCIAL
 0.0335 9 LUGAR_PROCEDENCIA
 0.0161 12 ESTADO_CIVIL

Apéndice G. Modelo J48

SANCION = NO

| ESTADO_ALUMNO = INACTIVO
 | | SEMESTRE_ACTUAL = PRIMERO
 | | | EDAD = RG_30_31
 | | | | DPTO_PROCEDENCIA = OTRO: MUY_MALO (3.0/1.0)
 | | | | DPTO_PROCEDENCIA = CESAR: MUY_MALO (2.0/1.0)
 | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MEDIO_BAJO (1.0)
 | | | EDAD = RG_34_35: MUY_MALO (2.0/1.0)
 | | | EDAD = RG_28_29
 | | | | VALORACION_ICFES = MEDIO: MEDIO (4.0/2.0)
 | | | | VALORACION_ICFES = MEDIO_BAJO
 | | | | | ESTRATO_SOCIAL = DOS: BAJO (2.0/1.0)
 | | | | | ESTRATO_SOCIAL = UNO: MUY_MALO (1.0)
 | | | | | ESTRATO_SOCIAL = CUATRO: MUY_MALO (0.0)
 | | | | | ESTRATO_SOCIAL = TRES: MUY_MALO (3.0/2.0)
 | | | | | ESTRATO_SOCIAL = SEIS: MUY_MALO (0.0)
 | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO_ALTO (1.0)
 | | | | VALORACION_ICFES = ALTO: MEDIO (0.0)
 | | | | VALORACION_ICFES = BAJO: MEDIO (0.0)
 | | | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
 | | | EDAD = RG_32_33: MUY_MALO (7.0/1.0)
 | | | EDAD = RG_36_37: MUY_MALO (1.0)
 | | | EDAD = RG_38_39: MUY_MALO (1.0)
 | | | EDAD = RG_40_adelante: MEDIO (1.0)

```

| | | EDAD = RG_26_27: MEDIO (12.0/3.0)
| | | EDAD = RG_22_23
| | | ANO_INGRESO = RG_2006_2010: MEDIO (0.0)
| | | ANO_INGRESO = RG_2011_2015
| | | | VALORACION_ICFES = MEDIO
| | | | | ESTRATO_SOCIAL = DOS: MUY_MALO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = UNO
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO (6.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO_ALTO (1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MEDIO (2.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | | | | ESTRATO_SOCIAL = CUATRO: MEDIO (0.0)
| | | | | | ESTRATO_SOCIAL = TRES: MEDIO (1.0)
| | | | | | ESTRATO_SOCIAL = SEIS: MEDIO (0.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO
| | | | | | ESTRATO_SOCIAL = DOS: MEDIO (5.0/1.0)
| | | | | | ESTRATO_SOCIAL = UNO
| | | | | | | SEXO = M: MUY_MALO (3.0/1.0)
| | | | | | | SEXO = F: MEDIO (2.0)
| | | | | | | ESTRATO_SOCIAL = CUATRO: MEDIO (0.0)
| | | | | | | ESTRATO_SOCIAL = TRES: MEDIO (0.0)
| | | | | | | ESTRATO_SOCIAL = SEIS: MEDIO (0.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO_ALTO (2.0)
| | | | | VALORACION_ICFES = ALTO: MEDIO (0.0)
| | | | | VALORACION_ICFES = BAJO: MEDIO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
| | | | | ANO_INGRESO = RG_2016_2020: MUY_MALO (5.0)
| | | EDAD = RG_24_25
| | | | ESTRATO_SOCIAL = DOS: MUY_BAJO (4.0/3.0)
| | | | | ESTRATO_SOCIAL = UNO
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO (7.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MUY_MALO (3.0/2.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MEDIO (8.0/3.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (2.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO_ALTO (1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | | | | ESTRATO_SOCIAL = CUATRO: MEDIO (0.0)
| | | | | | ESTRATO_SOCIAL = TRES: MEDIO_BAJO (1.0)
| | | | | | ESTRATO_SOCIAL = SEIS: MEDIO (0.0)
| | | EDAD = RG_20_21
| | | | VALORACION_ICFES = MEDIO: MUY_MALO (4.0/1.0)
| | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO (1.0)
| | | | VALORACION_ICFES = MEDIO_ALTO
| | | | | DPTO_PROCEDENCIA = OTRO: MEDIO (1.0)
| | | | | DPTO_PROCEDENCIA = CESAR: MUY_MALO (3.0/1.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER
| | | | | | ESTRATO_SOCIAL = DOS
| | | | | | | SEXO = M: MEDIO_ALTO (2.0)
| | | | | | | SEXO = F: MEDIO (2.0)
| | | | | | | ESTRATO_SOCIAL = UNO: MEDIO_ALTO (3.0)
| | | | | | | ESTRATO_SOCIAL = CUATRO: MEDIO (1.0)
| | | | | | | ESTRATO_SOCIAL = TRES: MEDIO_ALTO (0.0)
| | | | | | | ESTRATO_SOCIAL = SEIS: MEDIO_ALTO (0.0)
| | | | | VALORACION_ICFES = ALTO: MUY_MALO (0.0)
| | | | | VALORACION_ICFES = BAJO: MUY_MALO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: MUY_MALO (1.0)
| | | EDAD = RG_18_19: MEDIO (1.0)
| | | EDAD = RG_16_17: MEDIO (0.0)
| | | SEMESTRE_ACTUAL = SEGUNDO
| | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO (36.0/4.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = DOS
| | | | | SEXO = M: MEDIO (7.0/2.0)

```

```

| | | SEXO = F: MEDIO_ALTO (3.0/1.0)
| | | NUMERO_HERMANOS_ESTUDIAN = UNO
| | | EDAD = RG_30_31: MEDIO (0.0)
| | | EDAD = RG_34_35: MEDIO (0.0)
| | | EDAD = RG_28_29: MEDIO (0.0)
| | | EDAD = RG_32_33: MEDIO (0.0)
| | | EDAD = RG_36_37: MEDIO (0.0)
| | | EDAD = RG_38_39: MEDIO (0.0)
| | | EDAD = RG_40_adelante: MEDIO (0.0)
| | | EDAD = RG_26_27: MEDIO_ALTO (1.0)
| | | EDAD = RG_22_23: MEDIO (4.0/1.0)
| | | EDAD = RG_24_25: MEDIO (4.0)
| | | EDAD = RG_20_21
| | | | ANO_INGRESO = RG_2006_2010: MEDIO_ALTO (0.0)
| | | | ANO_INGRESO = RG_2011_2015: MEDIO_ALTO (2.0)
| | | | ANO_INGRESO = RG_2016_2020: MEDIO (3.0/1.0)
| | | EDAD = RG_18_19: MEDIO (3.0/1.0)
| | | EDAD = RG_16_17: MEDIO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO_ALTO (1.0)
| | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO_BAJO (2.0/1.0)
| | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (1.0)
| | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | SEMESTRE_ACTUAL = DECIMO
| | | VALORACION_ICFES = MEDIO
| | | | SEXO = M: MEDIO_ALTO (4.0/2.0)
| | | | SEXO = F: MEDIO (2.0)
| | | VALORACION_ICFES = MEDIO_BAJO: MEDIO (4.0)
| | | VALORACION_ICFES = MEDIO_ALTO: MEDIO (0.0)
| | | VALORACION_ICFES = ALTO: MEDIO (0.0)
| | | VALORACION_ICFES = BAJO: MEDIO (0.0)
| | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
| | | SEMESTRE_ACTUAL = OCTAVO: MEDIO (0.0)
| | | SEMESTRE_ACTUAL = TERCERO: MEDIO (18.0/1.0)
| | | SEMESTRE_ACTUAL = CUARTO: MEDIO (10.0)
| | | SEMESTRE_ACTUAL = SEXTO: MEDIO (5.0/1.0)
| | | SEMESTRE_ACTUAL = QUINTO: MEDIO (3.0/1.0)
| | | SEMESTRE_ACTUAL = NOVENO: MEDIO (1.0)
| | | SEMESTRE_ACTUAL = SEPTIMO: MEDIO_BAJO (1.0)
| | | ESTADO_ALUMNO = ACTIVO
| | | ANO_INGRESO = RG_2006_2010: MEDIO (4.0/1.0)
| | | ANO_INGRESO = RG_2011_2015
| | | VALORACION_ICFES = MEDIO
| | | | EDAD = RG_30_31: MEDIO (0.0)
| | | | EDAD = RG_34_35: MEDIO (0.0)
| | | | EDAD = RG_28_29: MEDIO (3.0)
| | | | EDAD = RG_32_33: MEDIO (1.0)
| | | | EDAD = RG_36_37: MEDIO (0.0)
| | | | EDAD = RG_38_39: MEDIO (0.0)
| | | | EDAD = RG_40_adelante: MEDIO (0.0)
| | | | EDAD = RG_26_27: MEDIO_ALTO (1.0)
| | | | EDAD = RG_22_23
| | | | DPTO_PROCEDENCIA = OTRO: MEDIO (3.0)
| | | | DPTO_PROCEDENCIA = CESAR: MEDIO_ALTO (4.0/1.0)
| | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER
| | | | | NUMERO_HERMANOS_ESTUDIAN = CERO
| | | | | SEMESTRE_ACTUAL = PRIMERO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = SEGUNDO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = DECIMO: MEDIO_ALTO (2.0)
| | | | | SEMESTRE_ACTUAL = OCTAVO: MEDIO (2.0)
| | | | | SEMESTRE_ACTUAL = TERCERO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = CUARTO: MEDIO (1.0)
| | | | | SEMESTRE_ACTUAL = SEXTO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = QUINTO: MEDIO (2.0/1.0)
| | | | | SEMESTRE_ACTUAL = NOVENO: MEDIO (2.0/1.0)
| | | | | SEMESTRE_ACTUAL = SEPTIMO: MEDIO (1.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO (2.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MEDIO (5.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (0.0)

```

```

| | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | | | EDAD = RG_24_25
| | | | | DPTO_PROCEDENCIA = OTRO: MEDIO_ALTO (2.0)
| | | | | DPTO_PROCEDENCIA = CESAR: MEDIO (3.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER
| | | | | NUMERO_HERMANOS_ESTUDIAN = CERO
| | | | | SEXO = M
| | | | | | ESTRATO_SOCIAL = DOS: MEDIO_ALTO (1.0)
| | | | | | ESTRATO_SOCIAL = UNO: MEDIO (3.0)
| | | | | | ESTRATO_SOCIAL = CUATRO: MEDIO (0.0)
| | | | | | ESTRATO_SOCIAL = TRES: MEDIO (2.0)
| | | | | | ESTRATO_SOCIAL = SEIS: MEDIO (0.0)
| | | | | | SEXO = F: MEDIO_ALTO (2.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO (2.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MEDIO (3.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO_ALTO (1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | | | EDAD = RG_20_21: MEDIO (4.0/2.0)
| | | | | EDAD = RG_18_19: MEDIO (0.0)
| | | | | EDAD = RG_16_17: MEDIO (0.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO (7.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO (4.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MEDIO_ALTO (4.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO
| | | | | | SEMESTRE_ACTUAL = PRIMERO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = SEGUNDO: MEDIO_ALTO (1.0)
| | | | | | SEMESTRE_ACTUAL = DECIMO: MEDIO_ALTO (4.0/1.0)
| | | | | | SEMESTRE_ACTUAL = OCTAVO: MEDIO (2.0/1.0)
| | | | | | SEMESTRE_ACTUAL = TERCERO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = CUARTO: MEDIO (2.0)
| | | | | | SEMESTRE_ACTUAL = SEXTO: MEDIO (2.0)
| | | | | | SEMESTRE_ACTUAL = QUINTO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = NOVENO: MEDIO_ALTO (6.0/1.0)
| | | | | | SEMESTRE_ACTUAL = SEPTIMO: MEDIO (5.0)
| | | | | VALORACION_ICFES = ALTO: MEDIO (0.0)
| | | | | VALORACION_ICFES = BAJO: MEDIO_ALTO (1.0)
| | | | | VALORACION_ICFES = MUY_ALTO: ALTO (1.0)
| | | | | ANO_INGRESO = RG_2016_2020
| | | | | | ESTRATO_SOCIAL = DOS
| | | | | | SEXO = M: MEDIO (27.0/1.0)
| | | | | | SEXO = F
| | | | | | | VALORACION_ICFES = MEDIO: MEDIO (3.0/1.0)
| | | | | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO (0.0)
| | | | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO (3.0)
| | | | | | | VALORACION_ICFES = ALTO: MEDIO_ALTO (1.0)
| | | | | | | VALORACION_ICFES = BAJO: MEDIO (0.0)
| | | | | | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
| | | | | | ESTRATO_SOCIAL = UNO
| | | | | | VALORACION_ICFES = MEDIO
| | | | | | | EDAD = RG_30_31: MEDIO (0.0)
| | | | | | | EDAD = RG_34_35: MEDIO (0.0)
| | | | | | | EDAD = RG_28_29: MEDIO (0.0)
| | | | | | | EDAD = RG_32_33: MEDIO (0.0)
| | | | | | | EDAD = RG_36_37: MEDIO (0.0)
| | | | | | | EDAD = RG_38_39: MEDIO (0.0)

```

```

| | | | EDAD = RG_40_adelante: MEDIO (0.0)
| | | | EDAD = RG_26_27: MEDIO (0.0)
| | | | EDAD = RG_22_23: MEDIO (7.0)
| | | | EDAD = RG_24_25: MEDIO (1.0)
| | | | EDAD = RG_20_21: MEDIO (14.0/1.0)
| | | | EDAD = RG_18_19
| | | | | SEXO = M: MEDIO (4.0)
| | | | | SEXO = F: MEDIO_ALTO (2.0)
| | | | EDAD = RG_16_17: MEDIO (0.0)
| | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO (1.0)
| | | | VALORACION_ICFES = MEDIO_ALTO
| | | | | EDAD = RG_30_31: MEDIO_ALTO (1.0)
| | | | | EDAD = RG_34_35: MEDIO (0.0)
| | | | | EDAD = RG_28_29: MEDIO (1.0)
| | | | | EDAD = RG_32_33: MEDIO (0.0)
| | | | | EDAD = RG_36_37: MEDIO (0.0)
| | | | | EDAD = RG_38_39: MEDIO (0.0)
| | | | | EDAD = RG_40_adelante: MEDIO (0.0)
| | | | | EDAD = RG_26_27: MEDIO (2.0)
| | | | | EDAD = RG_22_23: MEDIO (7.0/1.0)
| | | | | EDAD = RG_24_25: MEDIO (0.0)
| | | | | EDAD = RG_20_21
| | | | | DPTO_PROCEDENCIA = OTRO: MEDIO (4.0/2.0)
| | | | | DPTO_PROCEDENCIA = CESAR
| | | | | SEMESTRE_ACTUAL = PRIMERO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = SEGUNDO: MEDIO_ALTO (3.0/1.0)
| | | | | SEMESTRE_ACTUAL = DECIMO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = OCTAVO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = TERCERO: MEDIO (2.0/1.0)
| | | | | SEMESTRE_ACTUAL = CUARTO: MEDIO (1.0)
| | | | | SEMESTRE_ACTUAL = SEXTO: MEDIO (1.0)
| | | | | SEMESTRE_ACTUAL = QUINTO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = NOVENO: MEDIO (0.0)
| | | | | SEMESTRE_ACTUAL = SEPTIMO: MEDIO (0.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER
| | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO (8.0/1.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO (5.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = UNO
| | | | | | SEMESTRE_ACTUAL = PRIMERO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = SEGUNDO: MEDIO (4.0)
| | | | | | SEMESTRE_ACTUAL = DECIMO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = OCTAVO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = TERCERO: MEDIO (2.0)
| | | | | | SEMESTRE_ACTUAL = CUARTO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = SEXTO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = QUINTO: MEDIO_ALTO (1.0)
| | | | | | SEMESTRE_ACTUAL = NOVENO: MEDIO (0.0)
| | | | | | SEMESTRE_ACTUAL = SEPTIMO: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (2.0/1.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO (0.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO (0.0)
| | | | | EDAD = RG_18_19: MEDIO (21.0/1.0)
| | | | | EDAD = RG_16_17: MEDIO (0.0)
| | | | | VALORACION_ICFES = ALTO
| | | | | EDAD = RG_30_31: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_34_35: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_28_29: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_32_33: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_36_37: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_38_39: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_40_adelante: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_26_27: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_22_23: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_24_25: MEDIO_ALTO (0.0)
| | | | | EDAD = RG_20_21: MEDIO_ALTO (2.0)
| | | | | EDAD = RG_18_19: MEDIO (3.0/1.0)
| | | | | EDAD = RG_16_17: MEDIO_ALTO (0.0)

```

```

| | | VALORACION_ICFES = BAJO: MEDIO (0.0)
| | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
| | | ESTRATO_SOCIAL = CUATRO: MEDIO (0.0)
| | | ESTRATO_SOCIAL = TRES: MEDIO (2.0)
| | | ESTRATO_SOCIAL = SEIS: MEDIO (1.0)
| | | ESTADO_ALUMNO = EXCLUIDO: MEDIO (0.0)
SANCION = SI
| SEMESTRE_ACTUAL = PRIMERO
| | ESTADO_ALUMNO = INACTIVO
| | | NUMERO_HERMANOS_ESTUDIAN = CERO
| | | ANO_INGRESO = RG_2006_2010
| | | | EDAD = RG_30_31
| | | | | ESTRATO_SOCIAL = DOS
| | | | | | DPTO_PROCEDENCIA = OTRO: MUY_BAJO (1.0)
| | | | | | DPTO_PROCEDENCIA = CESAR
| | | | | | SEXO = M: BAJO (3.0/1.0)
| | | | | | SEXO = F: MEDIO_BAJO (3.0/1.0)
| | | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: BAJO (16.0/6.0)
| | | | | | ESTRATO_SOCIAL = UNO
| | | | | | SEXO = M
| | | | | | | VALORACION_ICFES = MEDIO: BAJO (8.0/2.0)
| | | | | | | VALORACION_ICFES = MEDIO_BAJO
| | | | | | | | DPTO_PROCEDENCIA = OTRO: MUY_BAJO (2.0/1.0)
| | | | | | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (1.0)
| | | | | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: BAJO (3.0/2.0)
| | | | | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (0.0)
| | | | | | | | VALORACION_ICFES = ALTO: BAJO (0.0)
| | | | | | | | VALORACION_ICFES = BAJO: BAJO (0.0)
| | | | | | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
| | | | | | | | SEXO = F
| | | | | | | | VALORACION_ICFES = MEDIO: MUY_MALO (2.0/1.0)
| | | | | | | | VALORACION_ICFES = MEDIO_BAJO: BAJO (2.0)
| | | | | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (0.0)
| | | | | | | | VALORACION_ICFES = ALTO: BAJO (0.0)
| | | | | | | | VALORACION_ICFES = BAJO: BAJO (0.0)
| | | | | | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
| | | | | | | | ESTRATO_SOCIAL = CUATRO: BAJO (0.0)
| | | | | | | | ESTRATO_SOCIAL = TRES: MUY_BAJO (6.0/2.0)
| | | | | | | | ESTRATO_SOCIAL = SEIS: BAJO (0.0)
| | | | | | | | EDAD = RG_34_35
| | | | | | | | ESTRATO_SOCIAL = DOS: BAJO (2.0/1.0)
| | | | | | | | ESTRATO_SOCIAL = UNO
| | | | | | | | | LUGAR_PROCEDENCIA = OTRO: MALO (2.0/1.0)
| | | | | | | | | LUGAR_PROCEDENCIA = OCANA: MUY_MALO (5.0/3.0)
| | | | | | | | | ESTRATO_SOCIAL = CUATRO: BAJO (2.0)
| | | | | | | | | ESTRATO_SOCIAL = TRES: BAJO (0.0)
| | | | | | | | | ESTRATO_SOCIAL = SEIS: BAJO (0.0)
| | | | | | | | | EDAD = RG_28_29
| | | | | | | | | LUGAR_PROCEDENCIA = OTRO
| | | | | | | | | SEXO = M
| | | | | | | | | | DPTO_PROCEDENCIA = OTRO: BAJO (2.0/1.0)
| | | | | | | | | | DPTO_PROCEDENCIA = CESAR
| | | | | | | | | | | VALORACION_ICFES = MEDIO: MEDIO_BAJO (6.0/3.0)
| | | | | | | | | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (2.0)
| | | | | | | | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (1.0)
| | | | | | | | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | | | | | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
| | | | | | | | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | | | | | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MUY_BAJO (6.0/2.0)
| | | | | | | | | | | SEXO = F
| | | | | | | | | | | VALORACION_ICFES = MEDIO: MEDIO_BAJO (3.0/1.0)
| | | | | | | | | | | VALORACION_ICFES = MEDIO_BAJO
| | | | | | | | | | | | ESTRATO_SOCIAL = DOS: MEDIO_BAJO (2.0)
| | | | | | | | | | | | ESTRATO_SOCIAL = UNO: BAJO (7.0/3.0)
| | | | | | | | | | | | ESTRATO_SOCIAL = CUATRO: MEDIO_BAJO (0.0)
| | | | | | | | | | | | ESTRATO_SOCIAL = TRES: MEDIO_BAJO (1.0)
| | | | | | | | | | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | | | | | | | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO_BAJO (0.0)
| | | | | | | | | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | | | | | | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)

```

```

| | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | | LUGAR_PROCEDENCIA = OCANA
| | | | | ESTRATO_SOCIAL = DOS: BAJO (15.0/8.0)
| | | | | ESTRATO_SOCIAL = UNO
| | | | | VALORACION_ICFES = MEDIO
| | | | | SEXO = M: MEDIO_BAJO (3.0/1.0)
| | | | | SEXO = F: MALO (3.0/1.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO: BAJO (12.0/5.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (0.0)
| | | | | VALORACION_ICFES = ALTO: BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
| | | | | ESTRATO_SOCIAL = CUATRO: BAJO (0.0)
| | | | | ESTRATO_SOCIAL = TRES: MUY_MALO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = SEIS: BAJO (0.0)
| | | | | EDAD = RG_32_33
| | | | | ESTRATO_SOCIAL = DOS: MEDIO_BAJO (15.0/8.0)
| | | | | ESTRATO_SOCIAL = UNO
| | | | | SEXO = M
| | | | | DPTO_PROCEDENCIA = OTRO: MUY_MALO (2.0/1.0)
| | | | | DPTO_PROCEDENCIA = CESAR: MUY_BAJO (1.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MALO (2.0/1.0)
| | | | | SEXO = F
| | | | | VALORACION_ICFES = MEDIO: BAJO (2.0/1.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO: BAJO (2.0/1.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: MUY_MALO (1.0)
| | | | | VALORACION_ICFES = ALTO: BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
| | | | | ESTRATO_SOCIAL = CUATRO: MALO (1.0)
| | | | | ESTRATO_SOCIAL = TRES: BAJO (2.0)
| | | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_36_37: BAJO (3.0/1.0)
| | | | | EDAD = RG_38_39: BAJO (0.0)
| | | | | EDAD = RG_40_adelante: MUY_MALO (1.0)
| | | | | EDAD = RG_26_27
| | | | | VALORACION_ICFES = MEDIO
| | | | | ESTRATO_SOCIAL = DOS: MEDIO_BAJO (4.0/1.0)
| | | | | ESTRATO_SOCIAL = UNO: MEDIO_BAJO (9.0)
| | | | | ESTRATO_SOCIAL = CUATRO: MEDIO_BAJO (0.0)
| | | | | ESTRATO_SOCIAL = TRES: MALO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO
| | | | | SEXO = M
| | | | | ESTRATO_SOCIAL = DOS: BAJO (3.0/1.0)
| | | | | ESTRATO_SOCIAL = UNO: BAJO (7.0/2.0)
| | | | | ESTRATO_SOCIAL = CUATRO: BAJO (0.0)
| | | | | ESTRATO_SOCIAL = TRES
| | | | | LUGAR_PROCEDENCIA = OTRO: MUY_BAJO (2.0)
| | | | | LUGAR_PROCEDENCIA = OCANA: BAJO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = SEIS: BAJO (0.0)
| | | | | SEXO = F
| | | | | LUGAR_PROCEDENCIA = OTRO: MEDIO_BAJO (5.0/3.0)
| | | | | LUGAR_PROCEDENCIA = OCANA
| | | | | ESTRATO_SOCIAL = DOS: BAJO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = UNO: MALO (3.0/1.0)
| | | | | ESTRATO_SOCIAL = CUATRO: BAJO (0.0)
| | | | | ESTRATO_SOCIAL = TRES: BAJO (0.0)
| | | | | ESTRATO_SOCIAL = SEIS: BAJO (0.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO_BAJO (1.0)
| | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_22_23: BAJO (0.0)
| | | | | EDAD = RG_24_25: BAJO (0.0)
| | | | | EDAD = RG_20_21: BAJO (0.0)
| | | | | EDAD = RG_18_19: BAJO (0.0)
| | | | | EDAD = RG_16_17: BAJO (0.0)
| | | | | ANO_INGRESO = RG_2011_2015
| | | | | EDAD = RG_30_31: BAJO (3.0/1.0)

```

```

| | | | | EDAD = RG_34_35: BAJO (2.0)
| | | | | EDAD = RG_28_29
| | | | | ESTRATO_SOCIAL = DOS: MUY_MALO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = UNO
| | | | | VALORACION_ICFES = MEDIO
| | | | | DPTO_PROCEDENCIA = OTRO: MEDIO_BAJO (1.0)
| | | | | DPTO_PROCEDENCIA = CESAR: MUY_MALO (2.0/1.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MUY_MALO (2.0/1.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (4.0/1.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO_BAJO (1.0)
| | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | | ESTRATO_SOCIAL = CUATRO: MEDIO_BAJO (0.0)
| | | | | ESTRATO_SOCIAL = TRES: MALO (1.0)
| | | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_32_33: MUY_MALO (1.0)
| | | | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_40_adelante: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_26_27
| | | | | VALORACION_ICFES = MEDIO
| | | | | LUGAR_PROCEDENCIA = OTRO
| | | | | SEXO = M: MEDIO_BAJO (6.0/2.0)
| | | | | SEXO = F: MUY_MALO (3.0/1.0)
| | | | | LUGAR_PROCEDENCIA = OCANA: BAJO (4.0/2.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO
| | | | | SEXO = M: BAJO (5.0/2.0)
| | | | | SEXO = F
| | | | | LUGAR_PROCEDENCIA = OTRO: MEDIO_BAJO (2.0)
| | | | | LUGAR_PROCEDENCIA = OCANA: MALO (3.0/2.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_22_23
| | | | | VALORACION_ICFES = MEDIO: MEDIO_BAJO (7.0/2.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (3.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: MUY_BAJO (1.0)
| | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_24_25
| | | | | ESTRATO_SOCIAL = DOS: MEDIO_BAJO (9.0/1.0)
| | | | | ESTRATO_SOCIAL = UNO
| | | | | SEXO = M: MEDIO_BAJO (19.0/11.0)
| | | | | SEXO = F
| | | | | DPTO_PROCEDENCIA = OTRO: BAJO (2.0/1.0)
| | | | | DPTO_PROCEDENCIA = CESAR: BAJO (8.0/5.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MUY_BAJO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = CUATRO: MEDIO_BAJO (0.0)
| | | | | ESTRATO_SOCIAL = TRES: MUY_MALO (2.0/1.0)
| | | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_20_21: MEDIO_BAJO (2.0)
| | | | | EDAD = RG_18_19: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | | | ANO_INGRESO = RG_2016_2020: MEDIO_BAJO (5.0)
| | | | | NUMERO_HERMANOS_ESTUDIAN = DOS
| | | | | EDAD = RG_30_31: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_34_35: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_28_29: MEDIO (1.0)
| | | | | EDAD = RG_32_33: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_40_adelante: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_26_27
| | | | | VALORACION_ICFES = MEDIO: MEDIO_BAJO (2.0/1.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO: BAJO (2.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (2.0/1.0)
| | | | | VALORACION_ICFES = ALTO: BAJO (0.0)

```

```

| | | | VALORACION_ICFES = BAJO: BAJO (0.0)
| | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
| | | | EDAD = RG_22_23
| | | | DPTO_PROCEDENCIA = OTRO
| | | | | VALORACION_ICFES = MEDIO: BAJO (2.0)
| | | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (2.0)
| | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (0.0)
| | | | | VALORACION_ICFES = ALTO: BAJO (0.0)
| | | | | VALORACION_ICFES = BAJO: BAJO (0.0)
| | | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
| | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (3.0)
| | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MEDIO_BAJO (4.0/1.0)
| | | | EDAD = RG_24_25
| | | | | DPTO_PROCEDENCIA = OTRO: MEDIO_BAJO (2.0)
| | | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (3.0/1.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: BAJO (4.0/1.0)
| | | | EDAD = RG_20_21: MUY_BAJO (2.0/1.0)
| | | | EDAD = RG_18_19: MEDIO_BAJO (3.0)
| | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = UNO
| | | | | EDAD = RG_30_31: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_34_35: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_28_29: MEDIO_BAJO (1.0)
| | | | | EDAD = RG_32_33: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_40_adelante: MUY_MALO (1.0)
| | | | | EDAD = RG_26_27: BAJO (4.0/3.0)
| | | | | EDAD = RG_22_23: MEDIO_BAJO (11.0/2.0)
| | | | | EDAD = RG_24_25: MEDIO_BAJO (11.0/2.0)
| | | | | EDAD = RG_20_21: MEDIO_BAJO (9.0)
| | | | | EDAD = RG_18_19: MEDIO_BAJO (1.0)
| | | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO_BAJO (5.0/1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO_BAJO (4.0/2.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO_BAJO (1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO_BAJO (0.0)
| | | | ESTADO_ALUMNO = ACTIVO
| | | | | EDAD = RG_30_31: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_34_35: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_28_29: MEDIO (1.0)
| | | | | EDAD = RG_32_33: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_40_adelante: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_26_27: MEDIO_BAJO (0.0)
| | | | | EDAD = RG_22_23: MEDIO_BAJO (3.0/1.0)
| | | | | EDAD = RG_24_25: MEDIO_BAJO (1.0)
| | | | | EDAD = RG_20_21: MEDIO_BAJO (5.0)
| | | | | EDAD = RG_18_19: MUY_MALO (2.0/1.0)
| | | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | | ESTADO_ALUMNO = EXCLUIDO
| | | | | ESTRATO_SOCIAL = DOS
| | | | | DPTO_PROCEDENCIA = OTRO: BAJO (4.0)
| | | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (9.0/2.0)
| | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER
| | | | | ANO_INGRESO = RG_2006_2010: BAJO (0.0)
| | | | | ANO_INGRESO = RG_2011_2015
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: BAJO (4.0/2.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: BAJO (3.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: BAJO (4.0/1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MUY_BAJO (1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: BAJO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: BAJO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: BAJO (1.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: BAJO (0.0)
| | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: BAJO (0.0)
| | | | | ANO_INGRESO = RG_2016_2020

```

| | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO_BAJO (3.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MUY_BAJO (1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MALO (3.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MALO (0.0)
 | | | | | ESTRATO_SOCIAL = UNO
 | | | | | EDAD = RG_30_31: MUY_MALO (1.0)
 | | | | | EDAD = RG_34_35: BAJO (0.0)
 | | | | | EDAD = RG_28_29: MUY_BAJO (3.0/1.0)
 | | | | | EDAD = RG_32_33: BAJO (0.0)
 | | | | | EDAD = RG_36_37: BAJO (0.0)
 | | | | | EDAD = RG_38_39: BAJO (0.0)
 | | | | | EDAD = RG_40_adelante: BAJO (0.0)
 | | | | | EDAD = RG_26_27: MALO (8.0/4.0)
 | | | | | EDAD = RG_22_23
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO
 | | | | | VALORACION_ICFES = MEDIO: BAJO (5.0/3.0)
 | | | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (2.0)
 | | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (1.0)
 | | | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
 | | | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
 | | | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS
 | | | | | VALORACION_ICFES = MEDIO: BAJO (2.0)
 | | | | | VALORACION_ICFES = MEDIO_BAJO: MUY_BAJO (2.0/1.0)
 | | | | | VALORACION_ICFES = MEDIO_ALTO: BAJO (2.0/1.0)
 | | | | | VALORACION_ICFES = ALTO: BAJO (0.0)
 | | | | | VALORACION_ICFES = BAJO: BAJO (0.0)
 | | | | | VALORACION_ICFES = MUY_ALTO: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO
 | | | | | LUGAR_PROCEDENCIA = OTRO: BAJO (5.0/2.0)
 | | | | | LUGAR_PROCEDENCIA = OCANA: MEDIO_BAJO (4.0/2.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: BAJO (3.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MUY_BAJO (2.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: BAJO (0.0)
 | | | | | EDAD = RG_24_25
 | | | | | VALORACION_ICFES = MEDIO
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MUY_MALO (2.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MALO (2.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MUY_MALO (2.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MUY_BAJO (1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MUY_MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MUY_MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MUY_MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MUY_MALO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MUY_MALO (0.0)
 | | | | | VALORACION_ICFES = MEDIO_BAJO
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CERO
 | | | | | DPTO_PROCEDENCIA = OTRO: BAJO (2.0/1.0)
 | | | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (4.0/2.0)
 | | | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER
 | | | | | LUGAR_PROCEDENCIA = OTRO: MALO (3.0/1.0)
 | | | | | LUGAR_PROCEDENCIA = OCANA: MUY_BAJO (2.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MUY_MALO (2.0/1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = UNO: BAJO (4.0/2.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = TRES: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: BAJO (1.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: BAJO (0.0)
 | | | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: BAJO (0.0)
 | | | | | VALORACION_ICFES = MEDIO_ALTO: MALO (2.0)
 | | | | | VALORACION_ICFES = ALTO: BAJO (0.0)

```

| | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | | ANO_INGRESO = RG_2016_2020
| | | | DPTO_PROCEDENCIA = OTRO: MEDIO_BAJO (0.0)
| | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (2.0)
| | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: MEDIO (3.0/1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = DOS
| | | | EDAD = RG_30_31: MEDIO_BAJO (0.0)
| | | | EDAD = RG_34_35: MEDIO_BAJO (0.0)
| | | | EDAD = RG_28_29: MEDIO_BAJO (0.0)
| | | | EDAD = RG_32_33: MEDIO_BAJO (0.0)
| | | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | | EDAD = RG_40_adelante: MEDIO_BAJO (0.0)
| | | | EDAD = RG_26_27: MEDIO (2.0)
| | | | EDAD = RG_22_23: MEDIO_BAJO (3.0/1.0)
| | | | EDAD = RG_24_25: MEDIO_BAJO (4.0/2.0)
| | | | EDAD = RG_20_21: MEDIO_BAJO (4.0/1.0)
| | | | EDAD = RG_18_19: MEDIO_BAJO (1.0)
| | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = UNO
| | | | LUGAR_PROCEDENCIA = OTRO
| | | | SEXO = M
| | | | DPTO_PROCEDENCIA = OTRO: MEDIO_BAJO (2.0/1.0)
| | | | DPTO_PROCEDENCIA = CESAR: BAJO (3.0/1.0)
| | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: BAJO (1.0)
| | | | SEXO = F
| | | | DPTO_PROCEDENCIA = OTRO: MEDIO_BAJO (0.0)
| | | | DPTO_PROCEDENCIA = CESAR: MEDIO_BAJO (4.0)
| | | | DPTO_PROCEDENCIA = NORTE_DE_SANTANDER: BAJO (2.0/1.0)
| | | | LUGAR_PROCEDENCIA = OCANA: MEDIO_BAJO (9.0/1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO (3.0/1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO_BAJO (0.0)
| | | | ESTADO_ALUMNO = ACTIVO
| | | | ANO_INGRESO = RG_2006_2010: MEDIO_BAJO (0.0)
| | | | ANO_INGRESO = RG_2011_2015: MEDIO_BAJO (8.0/1.0)
| | | | ANO_INGRESO = RG_2016_2020
| | | | VALORACION_ICFES = MEDIO: MEDIO_BAJO (13.0/1.0)
| | | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (2.0)
| | | | VALORACION_ICFES = MEDIO_ALTO
| | | | NUMERO_HERMANOS_ESTUDIAN = CERO: MEDIO (4.0/1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO (5.0/2.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = UNO: MEDIO_BAJO (2.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO_BAJO (0.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO_BAJO (1.0)
| | | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO_BAJO (0.0)
| | | | VALORACION_ICFES = ALTO: MEDIO_BAJO (0.0)
| | | | VALORACION_ICFES = BAJO: MEDIO_BAJO (0.0)
| | | | VALORACION_ICFES = MUY_ALTO: MEDIO_BAJO (0.0)
| | | | ESTADO_ALUMNO = EXCLUIDO: MEDIO_BAJO (9.0/4.0)
| | | | SEMESTRE_ACTUAL = DECIMO
| | | | ESTADO_ALUMNO = INACTIVO: MEDIO_BAJO (5.0/2.0)
| | | | ESTADO_ALUMNO = ACTIVO: MEDIO (2.0)
| | | | ESTADO_ALUMNO = EXCLUIDO: BAJO (3.0/1.0)
| | | | SEMESTRE_ACTUAL = OCTAVO
| | | | ANO_INGRESO = RG_2006_2010: BAJO (2.0/1.0)
| | | | ANO_INGRESO = RG_2011_2015: MEDIO (10.0)
| | | | ANO_INGRESO = RG_2016_2020: MEDIO (0.0)
| | | | SEMESTRE_ACTUAL = TERCERO
| | | | NUMERO_HERMANOS_ESTUDIAN = CERO
| | | | ESTADO_ALUMNO = INACTIVO
| | | | EDAD = RG_30_31: MEDIO_BAJO (6.0/2.0)
| | | | EDAD = RG_34_35: MEDIO_BAJO (1.0)
| | | | EDAD = RG_28_29: MEDIO_BAJO (3.0/1.0)

```

```

| | | EDAD = RG_32_33: MEDIO_BAJO (0.0)
| | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | EDAD = RG_40_adelante: MEDIO_BAJO (0.0)
| | | EDAD = RG_26_27
| | | SEXO = M: BAJO (2.0/1.0)
| | | SEXO = F
| | | | ESTRATO_SOCIAL = DOS: MEDIO (2.0)
| | | | ESTRATO_SOCIAL = UNO: MEDIO_BAJO (2.0)
| | | | ESTRATO_SOCIAL = CUATRO: MEDIO_BAJO (0.0)
| | | | ESTRATO_SOCIAL = TRES: MEDIO_BAJO (0.0)
| | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | EDAD = RG_22_23: BAJO (1.0)
| | | EDAD = RG_24_25: MEDIO_BAJO (2.0)
| | | EDAD = RG_20_21: MEDIO (1.0)
| | | EDAD = RG_18_19: MEDIO_BAJO (0.0)
| | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | ESTADO_ALUMNO = ACTIVO: MEDIO (6.0/1.0)
| | | ESTADO_ALUMNO = EXCLUIDO: MEDIO_BAJO (3.0)
| | | NUMERO_HERMANOS_ESTUDIAN = DOS: MEDIO_BAJO (5.0/1.0)
| | | NUMERO_HERMANOS_ESTUDIAN = UNO
| | | | EDAD = RG_30_31: MEDIO (1.0)
| | | | EDAD = RG_34_35: MEDIO_BAJO (0.0)
| | | | EDAD = RG_28_29: MEDIO_BAJO (0.0)
| | | | EDAD = RG_32_33: MEDIO_BAJO (0.0)
| | | | EDAD = RG_36_37: MEDIO_BAJO (0.0)
| | | | EDAD = RG_38_39: MEDIO_BAJO (0.0)
| | | | EDAD = RG_40_adelante: MEDIO_BAJO (0.0)
| | | | EDAD = RG_26_27: MEDIO_BAJO (1.0)
| | | | EDAD = RG_22_23: MEDIO_BAJO (4.0)
| | | | EDAD = RG_24_25: MEDIO (5.0/1.0)
| | | | EDAD = RG_20_21: MEDIO_BAJO (2.0/1.0)
| | | | EDAD = RG_18_19: MEDIO (1.0)
| | | | EDAD = RG_16_17: MEDIO_BAJO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = TRES: MEDIO_BAJO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = CUATRO: MEDIO_BAJO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = CINCO: MEDIO_BAJO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = SEIS: MEDIO_BAJO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = DIEZ: MEDIO_BAJO (0.0)
| | | NUMERO_HERMANOS_ESTUDIAN = SIETE: MEDIO_BAJO (0.0)
| | SEMESTRE_ACTUAL = CUARTO
| | | VALORACION_ICFES = MEDIO
| | | | ESTADO_ALUMNO = INACTIVO: MEDIO_BAJO (2.0/1.0)
| | | | ESTADO_ALUMNO = ACTIVO: MEDIO (9.0)
| | | | ESTADO_ALUMNO = EXCLUIDO: MEDIO_BAJO (3.0)
| | | VALORACION_ICFES = MEDIO_BAJO
| | | | ESTRATO_SOCIAL = DOS: MEDIO_BAJO (2.0)
| | | | ESTRATO_SOCIAL = UNO: MEDIO_BAJO (2.0)
| | | | ESTRATO_SOCIAL = CUATRO: MEDIO_BAJO (0.0)
| | | | ESTRATO_SOCIAL = TRES: MUY_MALO (1.0)
| | | | ESTRATO_SOCIAL = SEIS: MEDIO_BAJO (0.0)
| | | VALORACION_ICFES = MEDIO_ALTO: MEDIO (3.0)
| | | VALORACION_ICFES = ALTO: MEDIO (0.0)
| | | VALORACION_ICFES = BAJO: MEDIO (0.0)
| | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
| | SEMESTRE_ACTUAL = SEXTO
| | | VALORACION_ICFES = MEDIO: MEDIO (13.0/1.0)
| | | VALORACION_ICFES = MEDIO_BAJO: MEDIO_BAJO (4.0/1.0)
| | | VALORACION_ICFES = MEDIO_ALTO: MEDIO (2.0)
| | | VALORACION_ICFES = ALTO: MEDIO (0.0)
| | | VALORACION_ICFES = BAJO: MEDIO (0.0)
| | | VALORACION_ICFES = MUY_ALTO: MEDIO (0.0)
| | SEMESTRE_ACTUAL = QUINTO
| | | ESTADO_ALUMNO = INACTIVO: MEDIO_BAJO (4.0/1.0)
| | | ESTADO_ALUMNO = ACTIVO: MEDIO (10.0/1.0)
| | | ESTADO_ALUMNO = EXCLUIDO: MEDIO (0.0)
| | SEMESTRE_ACTUAL = NOVENO: MEDIO (3.0/1.0)
| | SEMESTRE_ACTUAL = SEPTIMO: MEDIO (10.0)

```

Size of the tree : 863

Apéndice H. Modelo Naive Bayes

Attribute	Class								
	MUY_MALO (0.06)	BAJO (0.04)	MALO (0.06)	MUY_BAJO (0.29)	MEDIO_BAJO (0.34)	MEDIO (0.07)	MEDIO_ALTO (0.01)	ALTO	
=====									
ESTADO_ALUMNO									
INACTIVO	59.0	138.0	26.0	50.0	276.0	192.0	35.0	4.0	
ACTIVO	2.0	4.0	1.0	1.0	50.0	246.0	55.0	4.0	
EXCLUIDO	11.0	41.0	24.0	24.0	40.0	1.0	1.0	1.0	
[total]	72.0	183.0	51.0	75.0	366.0	439.0	91.0	9.0	
SANCION									
NO	32.0	4.0	1.0	4.0	11.0	322.0	88.0	7.0	
SI	39.0	178.0	49.0	70.0	354.0	116.0	2.0	1.0	
[total]	71.0	182.0	50.0	74.0	365.0	438.0	90.0	8.0	
ANO_INGRESO									
RG_2006_2010	21.0	87.0	19.0	34.0	115.0	52.0	5.0	2.0	
RG_2011_2015	41.0	84.0	24.0	34.0	172.0	212.0	59.0	5.0	
RG_2016_2020	10.0	12.0	8.0	7.0	79.0	175.0	27.0	2.0	
[total]	72.0	183.0	51.0	75.0	366.0	439.0	91.0	9.0	
SEMESTRE_ACTUAL									
PRIMERO	69.0	143.0	49.0	67.0	209.0	72.0	19.0	4.0	
SEGUNDO	1.0	30.0	1.0	5.0	108.0	134.0	21.0	2.0	
DECIMO	1.0	4.0	1.0	3.0	5.0	26.0	19.0	2.0	
OCTAVO	1.0	2.0	1.0	1.0	2.0	23.0	8.0	1.0	
TERCERO	1.0	6.0	1.0	1.0	26.0	52.0	7.0	1.0	
CUARTO	2.0	1.0	1.0	1.0	9.0	40.0	3.0	1.0	
SEXTO	1.0	1.0	1.0	1.0	6.0	34.0	2.0	1.0	
QUINTO	1.0	1.0	1.0	1.0	5.0	28.0	6.0	1.0	
NOVENO	1.0	1.0	1.0	1.0	1.0	13.0	10.0	2.0	
SEPTIMO	1.0	1.0	1.0	1.0	2.0	24.0	3.0	1.0	
[total]	79.0	190.0	58.0	82.0	373.0	446.0	98.0	16.0	
VALORACION_ICFES									
MEDIO	33.0	89.0	25.0	30.0	208.0	214.0	42.0	4.0	
MEDIO_BAJO	30.0	81.0	18.0	37.0	111.0	82.0	8.0	1.0	
MEDIO_ALTO	6.0	12.0	8.0	7.0	46.0	137.0	36.0	3.0	
ALTO	3.0	1.0	1.0	2.0	1.0	6.0	5.0	1.0	
BAJO	1.0	2.0	1.0	1.0	1.0	2.0	2.0	1.0	
MUY_ALTO	2.0	1.0	1.0	1.0	2.0	1.0	1.0	2.0	
[total]	75.0	186.0	54.0	78.0	369.0	442.0	94.0	12.0	
EDAD									
RG_30_31	6.0	31.0	3.0	13.0	36.0	26.0	4.0	1.0	
RG_34_35	4.0	7.0	2.0	4.0	5.0	4.0	2.0	1.0	
RG_28_29	9.0	38.0	10.0	13.0	51.0	32.0	6.0	1.0	
RG_32_33	13.0	11.0	5.0	3.0	13.0	6.0	1.0	2.0	
RG_36_37	2.0	3.0	1.0	1.0	3.0	1.0	2.0	1.0	
RG_38_39	2.0	1.0	1.0	1.0	2.0	2.0	1.0	1.0	
RG_40_adelante	3.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	

RG_26_27	7.0	29.0	13.0	14.0	52.0	40.0	5.0	2.0		
RG_22_23	11.0	30.0	5.0	11.0	71.0	99.0	25.0	1.0		
RG_24_25	14.0	33.0	13.0	16.0	69.0	86.0	19.0	1.0		
RG_20_21	8.0	5.0	4.0	5.0	56.0	94.0	27.0	4.0		
RG_18_19	2.0	3.0	2.0	2.0	16.0	55.0	7.0	2.0		
RG_16_17	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0		
[total]	82.0	193.0	61.0	85.0	376.0	449.0	101.0	19.0		
SEXO										
M	51.0	139.0	34.0	56.0	268.0	293.0	51.0	3.0		
F	20.0	43.0	16.0	18.0	97.0	145.0	39.0	5.0		
[total]	71.0	182.0	50.0	74.0	365.0	438.0	90.0	8.0		
ESTRATO_SOCIAL										
DOS	17.0	65.0	14.0	24.0	113.0	116.0	24.0	5.0		
UNO	48.0	102.0	30.0	42.0	227.0	302.0	61.0	3.0		
CUATRO	1.0	4.0	2.0	1.0	3.0	2.0	1.0	1.0		
TRES	7.0	13.0	6.0	9.0	24.0	19.0	6.0	1.0		
SEIS	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0		
[total]	74.0	185.0	53.0	77.0	368.0	441.0	93.0	11.0		
LUGAR_PROCEDENCIA										
OTRO	45.0	106.0	25.0	43.0	223.0	239.0	49.0	4.0		
OCANA	26.0	76.0	25.0	31.0	142.0	199.0	41.0	4.0		
[total]	71.0	182.0	50.0	74.0	365.0	438.0	90.0	8.0		
DPTO_PROCEDENCIA										
OTRO	15.0	29.0	6.0	14.0	55.0	66.0	13.0	1.0		
CESAR	24.0	50.0	13.0	18.0	134.0	112.0	19.0	2.0		
NORTE_DE_SANTANDER			33.0	104.0	32.0	43.0	177.0	261.0	59.0	6.0
[total]	72.0	183.0	51.0	75.0	366.0	439.0	91.0	9.0		
NUMERO_HERMANOS_ESTUDIAN										
CERO	50.0	133.0	38.0	51.0	223.0	214.0	36.0	2.0		
DOS	5.0	18.0	5.0	7.0	53.0	75.0	15.0	2.0		
UNO	16.0	23.0	7.0	12.0	74.0	120.0	30.0	4.0		
TRES	2.0	6.0	2.0	5.0	10.0	21.0	7.0	1.0		
CUATRO	1.0	3.0	1.0	2.0	5.0	6.0	4.0	2.0		
CINCO	1.0	1.0	1.0	1.0	2.0	2.0	2.0	1.0		
SEIS	1.0	3.0	1.0	1.0	2.0	4.0	1.0	1.0		
DIEZ	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0		
SIETE	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0		
[total]	78.0	189.0	57.0	81.0	372.0	445.0	97.0	15.0		

Apéndice I. Modelo RandomizableFilteredClassifier

```
@attribute K1 numeric
@attribute K2 numeric
@attribute K3 numeric
@attribute K4 numeric
@attribute K5 numeric
@attribute K6 numeric
@attribute K7 numeric
@attribute K8 numeric
@attribute K9 numeric
@attribute K10 numeric
@attribute ESTADO_PROMEDIO
{MUY_MALO,BAJO,MALO,MUY_BAJO,MEDIO_BAJO,MEDIO,MEDIO_ALTO,ALTO}
@data
```

Apéndice J. Manual de instalación

Este manual se usa para la creación de la base de datos y la instalación del sistema (Manual de instalación).

Apéndice K. Manual de usuario

Este manual es una guía acerca de la usabilidad del software para el usuario (Manual de usuario).