

**MINERIA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN
ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA
UNIVERSIDAD DE LA GUAJIRA**

AUTOR

DIEGO MADRID

**UNIVERSIDAD DE LA GUAJIRA
FACULTAD DE INGENIERIA
PROGRAMA: MAESTRIA EN GESTION DE LA TECNOLOGIA Y LA
INNOVACION
MAICAO – LA GUAJIRA
2019**

**MINERIA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN
ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA
UNIVERSIDAD DE LA GUAJIRA**

DIEGO MADRID

Propuesta presentada como requisito parcial para optar al título de Magíster
en Gestión de la Tecnología y la Innovación

DIRECTOR

HOBBER JOSE BERRIO CABALLERO
Dr. En Ciencias Gerenciales

UNIVERSIDAD DE LA GUAJIRA
FACULTAD DE INGENIERIA
PROGRAMA: MAESTRIA EN GESTION DE LA TECNOLOGIA Y LA
INNOVACION
MAICAO – LA GUAJIRA
2019

Diego Madrid. MINERIA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE LA GUAJIRA. Universidad de la Guajira. Facultad de Ingeniería. Programa: Maestría en Gestión de la Tecnología y la Innovación. Maicao – La Guajira. 2019.

RESUMEN

El objetivo de la investigación es evaluar el uso de la minería de datos para la estructura predictiva de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira. El mismo se fundamenta por los autores Song y Ying (2015); Intel IT Center (2014); González (2005); Acevedo y otros (2015); Grandon y Vargas (2012); Díaz (2008); Dult y otros (2017); Kovacic (2010); Eckert y Suenaga (2015); Mishra y otros (2014); Suarez-Tirado (2013); entre otros. En tanto, corresponde a una investigación de tipo descriptiva y proyectiva, de diseño no experimental, de campo, transaccional, con una población de 331 y una muestra de 181, con tres grupos de muestreo estudiantes, docentes y directiva, a la cual se le desarrollaron dos técnicas de recolección de datos la encuesta y la observación con 4 instrumentos de preguntas de selección y con escala de Likert, los mismos poseen 21, 12, 11 y 4 ítems. Asimismo, se construye un modelo de minería de retención para la predicción de la deserción estudiantil, empleando la técnica de minería: Árboles de decisión. Entre los resultados se presentan que los factores determinantes individuales y socioeconómicos no afectan la deserción estudiantil, pocos recursos en relación a los recursos tecnológicos, que el modelo de minería y recolección de información para predecir la deserción estudiantil no alcanzó variabilidad de resultados, además se observó que la influencia de la variable satisfacción con el programa se evidenció claramente en el valor de P chi-cuadrado corregido=0, 019, el cual resultó altamente significativo, lo cual nos dice que esa variable tuvo alta influencia en la deserción y se considera como la variable independiente con la mayor relación con la deserción y se recomienda aplicar el mismo modelo con una población mas alta para que las variables independiente se puedan evaluar con mas precision.

Palabras clave: Minería de datos, Predicción, Deserción estudiantil, Programa de Ingeniería en Sistemas.

INTRODUCCION

En la actualidad gran parte de las organizaciones, han descubierto un beneficio creciente al dotar gran parte de sus departamentos con tecnologías de punta y procesos competitivos y eficientes; para ello hace uso de los beneficios que ofrece la computación, las telecomunicaciones y la informática. Caso que se evidencia en las Instituciones Educativas de la República de Colombia. En este orden de ideas, esta ciencia aplicada hace referencia a la Minería de datos, la cual es considerada como una subdisciplina de las ciencias de la computación que presta apoyo a otras ciencias; su mayor fortaleza radica en la búsqueda de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles.

Asimismo, el Ministerio de Educación Nacional para la tasa de deserción en educación técnica y tecnológica consiste en reducir del 21.3%, porcentaje tomado como la línea de base en 2013, al 15% en 2018. Con estas disposiciones políticas vigentes y con el convencimiento de la utilidad de que un estudiante obtenga un título universitario, las instituciones de educación superior abordan la retención estudiantil desde sus proyectos educativos institucionales bajo sus propios criterios.

Por otro lado, la Minería de Datos, armada en procedimientos matemáticos de razonamiento, es viable llevar a cabo el análisis de amplios volúmenes de información recopilada en las bases de datos de las instituciones académicas; y manifestar modelos, agrupaciones, cambios o incoherencias, sobre datos que, tratados de forma encerrada, no contribuyen componentes que admitan tomar decisiones para investigar un caso que forma ruido, el cual se trata de las razones de deserción estudiantil.

En tal sentido, resultaría excelente equilibrar los perfiles del estudiante durante su constancia en la universidad y por ende el perfil del alumno desertor. Esto conlleva al gasto e inversiones de dinero y tiempo por lo cual dicho propietario busca la

mejor manera de obtener los recursos precisos tecnológicos y humanos para dichos desarrollos, entre unas de estas herramientas se tiene los prestamos como el más común en instituciones financieras o ayudas por parte del gobierno adecuadas para ello.

De igual forma, la Universidad de la Guajira posee una tasa de deserción por período según SPADIES (2017), en el 2014 evidencia una deserción de 8,09, en el 2015 un 29,42%, en el 2016 un 29,31% y los estudiantes admitidos en el año 2017 hasta la fecha tiene un 3,75%, dada a estas estadísticas y en relación a la problemática, a su vez, según estimaciones del Programa de Sistema de la Facultad de Ingeniería Sede Maicao, existe una deserción en el año 2017 de un 4,33%. Para ello, la investigación quedó estructurada de la siguiente manera:

En el capítulo I, se aborda lo relacionado al problema en estudio, sistematización, objetivos, justificación, y delimitación. El capítulo II, está conformado por la fundamentación teórica que sirvió de sustento a la investigación, detallando los antecedentes, fundamentos teóricos, marco conceptual, marco legal, marco institucional, hipótesis y sistema de variables. En el capítulo III se trata lo referente a los aspectos metodológicos utilizados para realizar el proyecto, partiendo del tipo y diseño de la investigación, población y muestra, líneas e instrumentos para recolectar datos, entre otros y, por último, en el capítulo IV se evidencia el análisis de los resultados, propuesta: modelo predictivo, lo cual genera conclusiones y recomendaciones.

CONTENIDO

	Pág.
1. EL PROBLEMA.....	1
1. Planteamiento del problema.....	1
1.1 Formulación del problema.....	7
1.1.1 Sistematización del Problema.....	7
1.2 Objetivos.....	8
1.2.1 Objetivo General.....	8
1.2.2. Objetivos Específicos.....	8
1.3 Justificación de la investigación.....	8
1.4 Delimitación de la investigación.....	10
2. MARCO DE REFERENCIA.....	12
2.1. MARCO TEÓRICO.....	12
2.1.1. Antecedentes Investigativos.....	12
2.1.2 Fundamentación teórica.....	22
2.1.2.1 Deserción estudiantil.....	22
2.1.2.1.1 Factores determinantes de la deserción.....	23
2.1.2.2 Minería de datos.....	32
2.1.2.3 Técnicas de Minería de Datos.....	33
2.1.2.4 Recursos necesarios.....	42
2.1.2.6 Modelos Predictivos.....	46
2.2. MARCO INSTITUCIONAL.....	50
2.3 SISTEMAS DE VARIABLES.....	58
3. MARCO METODOLOGICO.....	61
3.1 Enfoque.....	61
3.2 Tipo de investigación.....	62
3.3 Diseño de la investigación.....	64
3.4 Población y muestra.....	65
3.4.1 Población.....	65
3.4.2. Muestra.....	67
	68

3.5 Técnica de recolección de datos.....	70
3.6 Técnicas para el procesamiento de la información.....	74
4. RESULTADOS DE LA INVESTIGACIÓN.....	74
4.1 Análisis de los resultados.....	100
CONCLUSIONES.....	106
RECOMENDACIONES.....	107
REFERENCIAS BIBLIOGRAFICAS.....	120
ANEXOS.....	121
INSTRUMENTO A.....	124
INSTRUMENTO B.....	127
INSTRUMENTO C.....	129
INSTRUMENTO D.....	131

LISTA DE CUADROS

	Pág.
Cuadro 1. Matriz de Operacionalización de la Variable	61
Cuadro 2. Distribución de la población	67
Cuadro 3. Muestra	69

LISTA DE TABLAS

	Pág.
Tabla 1. Factores Individuales	76
Tabla 2. Factores Académicos.....	78
Tabla 3. Factores Académicos.....	82
Tabla 4. Factores Académicos.....	87
Tabla 5. Factores Institucionales	90
Tabla 6. Factores Socioeconómicos	93
Tabla 7. Recursos Necesarios	95
Tabla 8. Recuentos de la tabla cruzada. Piensa desertar del programa* Satisfacción con el programa	98
Tabla 9. Valores de los estadísticos y significancia de las pruebas.....	98
Tabla 10. Valor y significancia del Coeficiente de Contingencia.....	99
Tabla 11. Resumen del modelo de Árbol de Decisión obtenido	99
Tabla 12. Porcentajes de clasificación correcta.....	101

LISTA DE GRÁFICOS

	Pág.
Grafico 1. Frecuencia Relativa sobre los Factores Individuales.....	76
Grafico 2. Frecuencia Relativa sobre los factores académicos	78
Grafico 3. Frecuencia Relativa sobre los factores académicos	83
Grafico 4. Frecuencia Relativa sobre los factores académicos	87
Grafico 5. Frecuencia Relativa sobre los Factores Institucionales.....	90
Grafico 6. Frecuencia Relativa sobre los Factores Socioeconómicos	93
Grafico 7. Frecuencia Relativa sobre los Recursos Necesarios.....	95

LISTA DE FIGURAS

	Pág.
Figura 1. Representación gráfica del resultado obtenido por un algoritmo basado en la máquina de aprendizaje Árbol de Decisiones.	37
Figura 2. Representación gráfica de la segmentación según algoritmo de Árbol de decisión.....	38
Figura 3. Ejemplo de estructura de una red neuronal artificial	39
Figura 4. Representación gráfica de la aplicación del algoritmo SVM	42
Figura 5. Hiperplano separador en R2 SVM	42
Figura 6. Árbol de decisión para la Deserción del Programa	100

1. EL PROBLEMA

1.1 Planteamiento del Problema

Una de las características más importantes de la actualidad es la cantidad masiva de información que se fija en el espacio y que pasan por desapercibidas, es por ello que se llaman datos que los sabría definir como porciones pequeñas y características de un algo que de por sí solo no dan mucha referencia para que pudiesen servir como un ejemplo un dato es un color, como decir amarillo, pero al ser ligado con otro dato por ejemplo sexo femenino se podría decir que se trata de una mujer rubia.

Según las palabras mágicas de Francis Bacon el conocimiento es poder, esto ha obligado al crecimiento acelerado de toda institución a niveles insospechados, al tratamiento y recolección de todo tipo de información, por muy empírica que esta parezca ya que existen innumerables formas de cómo manipularlas. Aunque el principio de todo estudio es tener datos que poder tratar o manipular.

Aunado a ello, gran parte de las organizaciones, han descubierto un beneficio creciente al dotar gran parte de sus departamentos con tecnologías de punta y procesos competitivos y eficientes; para ello hace uso de los beneficios que ofrece la computación, las telecomunicaciones y la informática. Caso que se evidencia en las Instituciones Educativas de la República de Colombia. En este orden de ideas, esta ciencia aplicada hace referencia a la Minería de datos, la cual es considerada como una subdisciplina de las ciencias de la computación que presta apoyo a otras ciencias; su mayor fortaleza radica en la búsqueda de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles. (Valero 2009, p. 16).

Además, en año 2014, el Departamento Nacional de Planeación publicó que en Colombia se observó una disminución de 2.5 puntos porcentuales en la tasa de deserción de la educación superior, quedando en 10.4%. Adicional a ello, el

Departamento Nacional de Planeación instó a continuar las acciones que prevengan el abandono de la formación superior al ser un tema prioritario, sobre la base de las deserciones acontecidas en primer semestre para la educación técnica profesional del 33.8%, tecnológico del 25.8% y universitario del 18.47%.

Seguidamente, la meta planteada por el Ministerio de Educación Nacional para la tasa de deserción en educación técnica y tecnológica consiste en reducir del 21.3%, porcentaje tomado como la línea de base en 2013, al 15% en 2018. Con estas disposiciones políticas vigentes y con el convencimiento de la utilidad de que un estudiante obtenga un título universitario, las instituciones de educación superior abordan la retención estudiantil desde sus proyectos educativos institucionales bajo sus propios criterios y desde las directrices del Ministerio de Educación Nacional visibles en el Sistema de Prevención y Análisis de la Deserción en las Instituciones de Educación Superior Ministerio de Educación Nacional (2009).

Como complemento a lo descrito, el Banco Mundial, en el año 2017, señala que en el país Colombia, existen alrededor del 37% de los estudiantes que comienzan un programa universitario y luego abandonan el sistema de educación superior. Además, alrededor del 36% de los estudiantes que desertan en Colombia lo hacen al final del primer año, mientras que en Estados Unidos este porcentaje es del 15%.

Asimismo, que a través de la Minería de Datos, armada en procedimientos matemáticos de razonamiento, es viable llevar a cabo el análisis de amplios volúmenes de información recopilada en las bases de datos de las instituciones académicas; y manifestar modelos, agrupaciones, cambios o incoherencias, sobre datos que tratados de forma encerrada, no contribuyen componentes que admitan tomar decisiones para investigar un caso que forma ruido, el cual se trata de las razones de deserción estudiantil; procedidas por ejemplo del vacío estudio de los expedientes académicos del alumnado.

En tal sentido, resultaría excelente equilibrar los perfiles del estudiante durante su constancia en la universidad y por ende el perfil del alumno desertor. Esto conlleva al gasto e inversiones de dinero y tiempo por lo cual dicho propietario busca la mejor manera de obtener los recursos precisos tecnológicos y humanos para dichos desarrollos, entre unas de estas herramientas se tiene los prestamos como el más común en instituciones financieras o ayudas por parte del gobierno adecuadas para ello. Es así que cada día más y más empresas necesitan profesionales que puedan traducir las cantidades de datos existentes en valor para el negocio, de hecho, el científico de datos se ha propuesto como uno de los trabajos más atractivos para la próxima década. (Davenport y Patil, 2012, p. 75).

Desafortunadamente, a pesar de ésta creciente oportunidad, la demanda global de estos profesionales supera ampliamente la oferta de estas habilidades analíticas. (Manyika et al., 2011, p. 34); (Kelly et al., 2013, p. 25). Dado a ello, González (2005), detalla las consecuencias sociales que implica la deserción, entre las que se encuentran las expectativas de los estudiantes y sus familias y las consecuencias emocionales sobre los estudiantes.

Asimismo, la deserción estudiantil se considera un indicador importante para medir la movilidad social y reflejar la contribución social que realizan las universidades. (Grandón y Vargas, 2012, p. 27). En términos económicos, existe evidencia de que los estudiantes atribuyen su decisión de desertar de sus programas académicos a su situación económica. (Acevedo et al., 2015, p. 39).

Por ello, la deserción logra hallarse como el abandono de las asignaturas/cursos o los programas en los que se ha inscrito el estudiante, abandonando las clases y de practicar con los compromisos establecidos anticipadamente, lo cual posee derivaciones sobre los índices de la eficiencia posterior de una cohorte. Esta percepción ubica la deserción, como el abandono, entre los inconvenientes más complicados y habituales que afrontan las Instituciones de Educación Superior, de igual forma el atraso estudiantil y los bajos índices de eficiencia terminal.

Sumado al contexto, Anuiés (2007), ha descubierto que la deserción revela una combinación de factores que impresionan a los estudiantes. Entre ellos se encuentran: Las circunstancias económicas nocivas de los estudiantes, el incompleto nivel cultural de la familia al que alcanza, las perspectivas del estudiante con conexión a la importancia de la educación, la incompatibilidad del tiempo brindado al trabajo y a los estudios, las características personales del estudiante, el poco interés por los estudios en general, por la carrera y la institución.

Adicionalmente, las características precedentes del estudiante, como los bajos promedios derivados en la educación media que irradian la carencia de los conocimientos y las destrezas con que egresan los estudiantes, en relación con los citados, la deficiente disposición vocacional recibida, antes de integrarse a la educación universitaria, que induce que los alumnos se apunten en las carreras profesionales, sin respaldar su disposición en una consistente información sobre la misma.

Pese a la cantidad de los estudiantes que abandonan al comienzo de los estudios universitarios, casi el 30% de los que abandonan el sistema lo hacen después de cuatro años. Es por ello, que las tecnologías de minería de datos usan la potencia de la red distribuida de los recursos de computación y arquitectura que no comparte nada, marcos de proceso distribuido, y bases de datos no relacionales para redefinir la forma en que los datos se administran y analizan. Sin duda alguna, actualmente las bases de datos contienen una gran cantidad de datos susceptibles a ser “minables”, datos que muchas veces exceden las capacidades humanas de reducción y análisis, a fin de obtener información útil. (Kamber 2006, p. 44).

Por otro lado, Díaz (2008), señala que en las universidades latinoamericanas poseen recientes aumentos en cobertura que han generado bajas tasas de retención con la cual se ha estudiado el fenómeno de la deserción estudiantil tiene

relación al uso de minería de datos (MD) a través de modelos analíticos, ya sea descriptivos o predictivos.

Estos modelos descriptivos intentan encontrar patrones escondidos en los datos mientras que los modelos predictivos buscan predecir futuros comportamientos, entre ellos se encuentran modelos de retención, de acompañamiento y de duración. (Mishra et al., 2014, p. 33). En general, la minería de datos se utiliza tanto para extraer información útil desde grandes volúmenes de datos como para visualizarlos de una manera fácil de interpretar. (Song y Ying, 2015, p. 41).

Debido a ello, Han et al., (2011, p. 11), marcan que muchas personas tratan la minería de datos como un sinónimo de otro término utilizado popularmente, el descubrimiento de conocimiento de datos, o KDD, consiste en una secuencia iterativa de los siguientes pasos: 1. Datos de limpieza, integración de datos, la selección de datos, transformación de datos, la minería de datos, modelo de evaluación y la presentación del conocimiento, el cual es considerado como un modelo predictivo de duración de minería de datos para estimar la deserción estudiantil.

Adicional a lo plateado, la deserción estudiantil en los programas de pregrado de la gran mayoría de Instituciones de Educación Superior (IES) tanto de Colombia como de Latinoamérica es un problema que tiene un impacto muy amplio en el desarrollo social y económico de un país. América Latina afronta retos parecidos en el área educativa. La financiación, el aumento de la cobertura, el aseguramiento de la calidad, el mejoramiento de la igualdad en el acceso y permanencia, mayor articulación con la educación secundaria, multiplicidad de la oferta para atender distintas necesidades y campos de interés (ciencia, tecnología, investigación, humanidades, artes, formación integral) así como también mayor vinculación con el sector laboral y productivo (Timaran y Jiménez, 2014).

Por otro lado, Timaran et al. (2013), señalan que uno de los problemas que enfrenta el sistema de educación superior colombiano concierne a los altos niveles de deserción, pese a que los últimos años, según MEN (2009), se ha caracterizado por aumentos de cobertura e ingreso de estudiantes nuevos, el número de alumnos que logra culminar sus estudios superiores no es alto, dejando entrever que una gran parte entre estos abandona sus estudios, principalmente en los primeros semestres, ya que de cada cien estudiantes que ingresan en una institución de educación superior cerca de la mitad lo logra culminar su ciclo académico y obtener la graduación.

Asimismo, plantean que se estimó una deserción de 49% cuyas causas fueron: limitaciones económicas y financieras, bajo rendimiento académico, desorientación vocacional y profesional, además, dificultades para adaptarse al ambiente universitario, es por ello, se resalta que la deserción estudiantil conlleva a altos costos sociales y económicos que afectan a las familias, los estudiantes, las instituciones y el Estado.

Este resultado, fue adquirido según los datos de la Oficina de Control y Registro Académico (OCARA) de la Universidad de Nariño, en los programas de pregrado de la Universidad de Nariño, en las cohortes 2004-2006, donde ingresaron 6870 estudiantes, de los cuales, observados hasta el año 2011, desertaron 3366 estudiantes, correspondiente a un índice de deserción estudiantil del 49%.

De igual manera, en el periodo comprendido entre el primer semestre del 2004 y el segundo semestre de 2006, en la IUCESMAG ingresaron 1.054 estudiantes, de los cuales, hasta la ventana de observación del 2011, desertaron 589 estudiantes, correspondiente a un índice de deserción del 56%. El problema de deserción de estas dos IES se convirtió en un problema a resolver con minería de datos con el fin de detectar patrones de deserción estudiantil.

Añadido al planteamiento, la Universidad de la Guajira posee una tasa de deserción por período según SPADIES (2017), en el 2014 evidencia una deserción de 8,09, en el 2015 un 29,42%, en el 2016 un 29,31% y los estudiantes admitidos en el año 2017 hasta la fecha tiene un 3,75%, dada a estas estadísticas y en relación a la problemática, a su vez, según estimaciones del Programa de Sistema de la Facultad de Ingeniería Sede Maicao, existe una deserción en el año 2017 de un 4,33%.

En relación a lo planteado, se observa que en el Programa de Sistemas de la Facultad de Ingeniería de la Universidad de la Guajira estima que la deserción anual siempre ha estado por debajo de la media institucional, falta de un registro detallado de deserción por cohortes, desvíos del desempeño escolar, esto es, un bajo rendimiento, dificultades de conducta y aquellos relacionados a la edad, falta de interés de los estudiantes y familiares por la educación, dificultades familiares (Ama de casa, embarazo, etc.), inasistencia a clases (factores económicos), otras razones, tales como la discapacidad, servicio militar, enfermedad o accidente, asistencia a cursos especiales, desmotivación en los estudiantes, entre otros.

Además, la necesidad de obtener datos para identificar las causas de deserción de los estudiantes en los primeros semestres de la carrera de Ingeniería en Sistemas, es indispensable para tomar las acciones pertinentes y poder disminuir este índice, y no menos importante, predecir su deserción en cualquier momento, a efectos de su monitoreo y así poder tomar acciones correctivas.

Es por ello, que estos indicadores o factores están por debajo de su potencial, apenas la mitad de los estudiantes que ingresan a la educación superior obtiene su título entre los 25 y 29 años de edad, ya sea porque continúan estudiando o porque abandonaron los estudios; de no desarrollar controles para estos resultados, se prevé en un horizonte no muy lejano una reducción en la mano de obra calificada que afectaría la productividad del país. Con respecto a la Universidad de la Guajira Sede Maicao este indicador pondrá en desventaja la

sostenibilidad del programa académico de Ingeniería en Sistemas y la pérdida de la función social que la misma tiene.

En correspondencia con el planteamiento de esta investigación, se plantea a la minería de datos como aquel proceso mediante el cual se puede obtener un modelo que sirva para la predicción sobre el comportamiento de los datos, específicamente en lo relacionado a la deserción estudiantil. Este modelo se operacionaliza con el empleo o uso de bases de datos, aunado a la aplicación de algún algoritmo que construya el modelo.

Entre las propuestas que pretenden resolver algunas de estas problemáticas se incluye determinar los factores mediante la minería de datos y estructuras que faciliten la predicción de la deserción estudiantil en el programa de Ingeniería de Sistemas en la Universidad de la Guajira Sede Maicao, para así, optimizar la alta multiplicidad de información requerida en el análisis de la deserción estudiantil en dicho programa en el presente y a futuro.

1.1.1 Formulación del problema

Ante este contexto recorrido se hace autoritario el análisis de un sistema de un estudio orientado a:

¿Cómo el uso de la minería de datos afecta la estructura predictiva de la deserción estudiantil en el programa de Ingeniería de Sistemas en la Universidad de la Guajira?

1.1.2 Sistematización del problema

De la pregunta general se generan las siguientes interrogantes:

¿Cuáles son los factores determinantes de la deserción estudiantil en el programa de Ingeniería de Sistemas en la Universidad de la Guajira?

¿Qué recursos tecnológicos y humanos son necesarios para la aplicación de la minería de datos en la Universidad de la Guajira?

¿Cómo es el modelo predictivo en minería de datos que se ajuste a la dinámica de la deserción estudiantil?

1.2 Objetivos de la investigación

1.2.1 Objetivo general

Evaluar el uso de la minería de datos para la estructura predictiva de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.

1.2.2 Objetivos específicos

Caracterizar los factores determinantes de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.

Identificar los recursos necesarios para la aplicación de la minería de datos en la Universidad de la Guajira.

Describir el modelo predictivo en minería de datos que se ajuste a la dinámica de la deserción estudiantil.

Proponer mediante la minería de datos un modelo para la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.

1.3 Justificación de la investigación

Existe una necesidad latente y constante en mejorar la educación mundial como parte del desarrollo efectivo de la sociedad. La minería de datos se convierte en una herramienta que permite mejorar la calidad y pertinencia en la educación desde diferentes aspectos y con un beneficio para todos los actores del sistema educativo.

Es por ello, que la minería de datos aplicada a la educación se ha considerado como un tema de relevancia y tendencia, por lo cual tanto los gobiernos como las instituciones han comenzado a implementarlo a modo de estrategia tecnológica para el cumplimiento de su filosofía institucional. Indagar, analizar y generar oportunidades, es fundamental para los actores del sistema educativo colombiano con la intención de tener un mejor aprovechamiento del dato para luego evolucionar positivamente en una oferta educativa de calidad y pertinente.

La presente investigación pretende dar respuesta a argumentos observados en relación a la deserción estudiantil mediante la minería de datos en el Programa de Ingeniería en Sistemas en la Universidad de la Guajira. La misma se evidencia desde el punto de vista teórico dado a que la implementación de estrategias aplicadas al mejoramiento de la educación basadas en aplicación de minería de datos que se han encontrado son muy recientes por lo tanto aún no hay reportados resultados concretos; esto se debe considerar al momento de implementarlas en la ciudad ya que los impactos pueden verse reflejados más a hacia el mediano y largo plazo.

Esta estrategia tecnológica, contrarresta las consecuencias sociales que implica la deserción, entre las que se encuentran las expectativas de los estudiantes, sus familias y el fortalecimiento emocional que recae sobre ellos. Para lo expresado, el estudio es importante porque revela la realidad de la deserción estudiantil,

considerada un indicador importante para medir la movilidad social y reflejar la contribución social que realizan las universidades. (González, 2005, p. 28).

Asimismo, desde el punto de vista práctico es evidenciar de qué modo se presentan estructuras que faciliten la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira como estrategia del desarrollo de la educación universitaria, en busca de satisfacer las necesidades que tienen las instituciones, los procesos de almacenamiento de datos para la toma de decisiones, el personal y los coordinadores de programas de contar con un excelente conocimiento sobre el análisis de minería de datos, dado que cuando existe un seguimiento o un control en la información en todo su contenido, lo que conduce inspeccionar el resultado de la producción, la eficiencia y su eficacia.

En correspondencia al nivel metodológico la investigación busca lograr el desempeño de los objetivos del estudio a través del uso de técnicas de investigación, instrumentos para estructuras como estrategia del desarrollo de la educación, que perfeccionen la estimación de la deserción estudiantil de los diferentes programas, la operatividad, la productividad, la eficiencia y eficacia de las instituciones de educación superior o de otras instituciones educativas del Municipio Maicao y por ende el fortalecimiento del desarrollo educativo de las universidades. Asimismo, favorece a optimizar la variedad de la información generada en contextos de la deserción en el Programa de Ingeniería de Sistemas, la estructura que faciliten la predicción de la deserción estudiantil y sus requerimientos.

En cuanto a la estabilidad social, ayudar con la construcción de eficiente y eficaz información en el Programa de Ingeniería de Sistemas en la Universidad de la Guajira sede Maicao para que así se modernice y se fortalezca en el tiempo. Como también, para ser modelo de un fuerte sistema educativo y poder dar persistencia, reforzar el conocimiento, capacitar al personal en el área de minería

de datos en todos sus programas, lo cual conlleva a la institución a generar mayor productividad, equilibrio de recursos e ingresos, fortalecimiento del recurso técnico, económico y humano, como también del beneficio institucional y estudiantil para su productividad y crecimiento tecnológico.

1.4 Delimitación

El trabajo de investigación se considera viable por cuanto están garantizados los recursos financieros, los insumos y los medios para culminar con ella; se ajusta a la línea de investigación Gerencia Tecnológica, en la línea específica Gestión de la Tecnología y la Innovación. Se consideró como perspectivas teóricas a Song y Ying (2015); Intel IT Center (2014); González (2005); Acevedo y otros (2015); Grandon y Vargas (2012); Díaz (2008); Dult y otros (2017); Kovacic (2010); Eckert y Suenaga (2015); Mishra y otros (2014); Suarez-Tirado (2013); Philip y otros (2013); Sargut y Mcgrath (2011); Albarran y Salgado (2013); Davenport y Harris (2007); Chen y otros (2012); Kelly y otros (2013); entre otros.

Asimismo, se desarrollará en la Universidad la Guajira Sede Maicao, Programa de Ingeniería en Sistemas, la misma se le aplicará a la directiva, docentes y personal del dicho programa durante el período comprendido entre los meses de agosto de 2018 y octubre de 2019.

2. MARCO DE REFERENCIA

2.1 Marco teórico

2.1.1 Antecedentes de la investigación

Para llevar a cabo este estudio se investigan proyectos previos, que mediante su información y datos relevantes fortalecieron los aspectos teóricos que se indagaron sobre la imagen e identidad corporativa, cuya metodología y conclusiones interesaron de guía para lograr prácticas y discusiones que avale el objetivo de este estudio. En relación con las publicaciones previas realizadas, se tomó la investigación de:

El Centro de Estudios para el Desarrollo Económico de la Universidad de los Andes (CEDE) creó una herramienta informática que permite hacer el seguimiento al problema de la deserción en la educación Superior denominada Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior (SPADIES).

El cual según la información del MEN (2009), 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. De esta manera, permite 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 los estudiantes orientados a fomentar su permanencia y graduación.

Además, un informe presentado por centro de investigaciones y documentación socioeconómica (CIDSE, 2011) advierten que, en términos de la cuantificación del fenómeno, el SPADIES cumple con su objetivo, es una herramienta que además permite realizar un análisis descriptivo. Sin embargo, argumentan que, tanto desde

el punto de vista conceptual como metodológico, se presentan serias limitaciones en su capacidad predictiva y en especial cuando se tratan de hacerlas a nivel desagregado para cada institución.

Seguidamente, muestra que las estimaciones de los riesgos la obtienen utilizando la información suministrada, asumiendo que las IES no presentan diferencias entre ellas y que tampoco existe una diferenciación respecto a los programas académicos que se imparten en las instituciones. Tomando en cuenta esas advertencias cobra sentido hacer la búsqueda del mejor modelo estadístico que permita determinar los estudiantes que están en riesgo de desertar y que se ajuste a las características particulares de la IES que sería objeto de estudio y los factores identificados como determinantes en la deserción.

Por consiguiente, la UNAL en su plan Global de desarrollo 2016-2018 (UNAL 2015), presentó cifras estadísticas sobre el comportamiento de la deserción estudiantil para los programas de pregrado de cada una de sus Sedes, para ello utilizaron información extraída del SPADIES y emplearon el enfoque de medición por cohortes, la cual estima las probabilidades de deserción desde el primer semestre hasta el semestre quince, después de haber sido admitido un estudiante en la universidad.

1. Escenario Positivo (ideal): universidades, sedes de la universidad y programas académicos de pregrado que cuentan con una deserción definitiva acumulada al semestre menor al 50% y una deserción definitiva acumulada al primer semestre menor o igual al 20%.

2. Escenario semipositivo (no ideal): universidades, sedes de la universidad y programas académicos de pregrado que cuentan con una deserción definitiva acumulada al semestre menor o igual a un 50 % y una deserción definitiva acumulada a primer semestre mayor del 20%.

3. Escenario negativo (aplicar estrategias): universidades, sedes de la universidad y programas académicos de pregrado que cuentan con una deserción definitiva acumulada al semestre mayor a un 50 % y una deserción definitiva acumulada a primer semestre menor o igual al 20%.

4. Escenario negativo (para analizar): universidades, sedes de la universidad programas académicos de pregrado que cuentan con una deserción definitiva acumulada al semestre mayor a un 50% y una deserción definitiva acumulada a primer semestre mayor del 20%.

A su vez, la tasa histórica acumulada de deserción definitiva al primero y al semestre del país, de la UNAL y de cada una de sus Sedes Andinas (Bogotá, Medellín, Manizales y Palmira). En este punto llama la atención que la Sede Medellín es a única Sede de la UNAL que se encuentra ubicada en el escenario negativo y la que tiene una mayor deserción tanto en el primer semestre (16%) como en el 2do semestre (53%).

De igual forma, en la evaluación de la deserción por programas muestran que el 36% de los programas de pregrado presentan deserciones acumuladas al semestre superior a un 50%. Se presenta en el 14% de los programas de pregrado de la sede Bogotá, en el 70% de los programas de pregrado de la sede Medellín, en el 42% de los programas de la sede Manizales y en el 43%) de los programas de pregrado de la sede Palmira.

Asimismo, lo que implica que la Sede Medellín se presenta aproximadamente el 56% de la deserción a pesar de que solo tiene el 28% de los programas de pregrado la universidad. Tomando en cuenta estas cifras de deserción de la UNAL, el estudio se va a centrar en analizar los datos correspondientes a los estudiantes de la Sede Medellín, para lo cual se tomará la información de los estudiantes que ingresaron desde el primer semestre de 2009 hasta el segundo semestre de 2016.

Esta antecedente aportó referencias estadísticas para observar los niveles de deserción en diferentes programas en la educación superior. Lo cual generó aportes para la justificación de la investigación, como teoría de la deserción estudiantil por autores avalados, garantizando así el eje de los fundamentos teóricos y posteriormente la construcción de conclusiones y recomendaciones, facilitando la discusión de una de las variables objeto de estudio, para así lograr por ende sus objetivos.

Por su parte, Rodríguez et al., (2016), elaboraron su artículo titulado, "Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos" en la Revista Científica, Vol. 12. N° 26. Bogotá, D.C. Existe un problema latente en la educación de nivel superior en Colombia, el cual tiene que ver con los altos índices de deserción académica. Esencialmente son muy pocas las estrategias que se han implementado con el fin de frenar la tasa de deserción, puesto que solo hasta el año 2003, se inician de manera formal los estudios para poder establecer cuáles son las condiciones que propician el abandono de los estudios.

Sin embargo, se desconocen las causas que conllevan a que un estudiante abandone su carrera, para ello en este artículo se hará uso de la Minería de Datos, por medio de la cual se pretende generar un modelo de Árbol de Decisión implementando el algoritmo J48 mediante el uso de la herramienta WEKA con el fin de poder identificar estas causas.

Sumado a lo expuesto, el artículo aportó elementos teóricos para la construcción de las bases teóricas en relación a las variables objeto de estudio minería de datos y deserción estudiantil, como una de las técnicas de la minería de datos llamada árbol de decisión, para así comprender su importancia, lo cual generará argumentos para las discusiones y conclusiones de la presente investigación.

Por otro lado, Amaya (2016), desarrolló un artículo el cual tiene por título "Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos en la Universidad Francisco de Paula Santander, Ocaña, Colombia". Barranquilla, Colombia. La minería de datos o Data Mining permite descubrir información oculta en grandes cantidades de datos, información que por procedimientos tradicionales es muy difícil de visualizar.

Esta rama de la computación permite manejar y clasificar grandes cantidades de datos, para lo cual se utilizan una gran variedad de técnicas, entre las que se encuentran los árboles de decisión C4.5 y el ID3 que han demostrado ser muy eficientes para casos específicos de predicción; este tipo de técnicas generan árboles que, de acuerdo a la complejidad del tema de estudio, pueden ser muy variables: se pueden obtener árboles con muchos nodos y hojas en el caso de ID3 y árboles más pequeños si se utiliza C4.5.

Este artículo muestra la construcción de un modelo predictivo de deserción estudiantil, caracterizando a los estudiantes de la Universidad Simón Bolívar con el objetivo de poder predecir la probabilidad de deserción de los estudiantes; dicho modelo demostró el desempeño de los algoritmos presentados para clasificar datos bajo contextos variables y la precisión de uno con respecto al otro. Para la creación del modelo se utilizó la herramienta WEKA que permite de forma muy eficiente el procesamiento y clasificación de los datos con resultados satisfactorios.

Los aportes más significativos de este artículo están asociados al conocimiento y uso de la diversidad de las técnicas de la minería de datos, lo cual ratifica la importancia de obtener fundamentos sobre las mismas y como son implementadas para la predicción de la deserción estudiantil una de las variables objeto de estudio como es en específico la técnica de árboles de decisión C4.5 y el ID3 que han demostrado ser muy eficientes para casos específicos de predicción, lo cual

genera mayor claridad para la construcción de los análisis de la presente investigación.

Además, Cerezo et al., (2015), titularon su estudio "Predicción del abandono universitario: variables Explicativas y medidas de prevención" en la Revista Fuentes N° 16 en la Universidad de Oviedo. El abandono de los estudios universitarios es un problema cuyos costes son altos tanto para el individuo como para la sociedad. Es por ello que la prevención del mismo es fundamental y cobra especial relevancia en el actual contexto de crisis económica.

Diversos autores han desarrollado investigaciones con el objetivo de establecer modelos predictivos de este fenómeno (Castaño, Gallón, Gómez y Vásquez, 2004; Trevizán, Beltrán y Cosolito, 2009; Goldenhersh, Coria y Saino, 2011; Sánchez, 2014). En este artículo se analizan dichos trabajos, identificando las ventajas y desventajas de las metodologías más utilizadas; análisis correlacionales, análisis de regresión logística, análisis de supervivencia y minería de datos.

La investigación cuyos resultados aquí se exponen, aplica la primera de las metodologías mencionadas, a fin de comprobar en lo que respecta al fenómeno del abandono el valor predictivo de las variables rendimiento académico previo, fecha de matriculación, rendimiento en primer curso de universidad y asistencia a clase.

Los resultados confirman la relación de dichas variables con el fenómeno estudiado. Dichos resultados son consistentes con los obtenidos por diversos autores a lo largo del tiempo, y en base a ellos se proponen dos tipos de medidas; por un lado, acciones encaminadas a facilitar el diagnóstico respecto al problema del abandono, y, por otro lado, medidas encaminadas a su prevención.

La presente investigación aportó elementos teóricos por diferentes autores patentados sobre modelos predictivos que se ajustan a la dinámica de la

deserción estudiantil una de las variables objeto de estudio, la cual se enfoca el maximizo nivel de importancia en la implementación de estos modelos, esto también complementa los fundamentos teóricos, para así utilizarlos y contrarrestarlos con las discusiones, conclusiones y recomendaciones de este estudio.

Como también, Barragan et al., (2017), desarrollaron su artículo titulado, "Acercamiento a la deserción estudiantil desde la integración social y académica" en la Revista de la Educación Superior Vol. 46. N° 02. En las políticas públicas colombianas relativas a la deserción estudiantil en la educación superior se solicita a las instituciones la formulación y el seguimiento de planes para mitigarla.

Para observar estas pautas gubernamentales, en este documento se propone la combinación novedosa de la teoría de grafos y los árboles de decisión bajo la idea de modelar la interacción espontánea de estudiantes en algunas redes sociales con miras a estimar la contribución de éstas en la integración social y académica. Se valida la propuesta con un caso de estudio, mostrando los algoritmos ejecutados mediante herramientas informáticas.

Este articulo aportó teorías sobre técnicas de minería de datos, la cual corresponde a una de las dimensiones de la presente investigación, la cual se describe a la técnica de árboles de decisión, la cual ha demostrado predecir en un alto porcentaje la deserción estudiantil, lo cual atribuye un ejemplo para la elaboración de la metodología en relación a la técnica a utilizar.

Asimismo, Merlino et al., (2013), titularon su artículo "Aplicación de un modelo de duración en programas de prevención de deserción universitaria REDIE, en la Revista Electrónica de Investigación Educativa, Vol. 15 N°. 3, en la Universidad Autónoma de Baja California Ensenada, México. Las prácticas institucionales vinculadas con la prevención de la deserción de una carrera universitaria

requieren cada vez más de instrumentos validados que permitan anticipar tal comportamiento.

En este sentido han demostrado utilidad decisiva distintos modelos estadísticos generados a partir de información referida a los propios estudiantes, sus hogares y su desempeño académico, entre otros determinantes. El objetivo principal es aplicar un modelo predictivo del riesgo de deserción de estudiantes universitarios a fin de generar resultados de manera temprana y progresivamente más eficaces.

El trabajo exhibe la utilidad de los modelos de duración en una muestra de estudiantes presenciales de la Universidad Autónoma de Baja California Ensenada, México, y la capacidad anticipatoria de los comportamientos de permanencia/deserción en el tiempo, a través de estimaciones de riesgo con un modelo de Cox en cuatro momentos de los primeros meses de universidad.

Este artículo aportó proposiciones sobre técnicas de minería de datos, la cual corresponde a una de las dimensiones de la presente investigación, la cual se describe a la técnica de duración en una muestra, la cual ha manifestado predecir la deserción estudiantil, lo cual imputa una muestra para la elaboración de los fundamentos teóricos en relación a los modelos predictivos para la deserción estudiantil y utilizarlos para las discusiones de la presente investigación.

Seguidamente, Matheus et al., (2018), titularon su artículo "Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins Santiago de Chile, Chile, en la Revista Educación Pesqui., São Paulo, Vol. 44. El presente estudio tiene como objetivo modelar un sistema predictivo de retención de los estudiantes de primer año de la Universidad Bernardo O'Higgins - UBO (Santiago de Chile), determinando cuáles de las variables de entrada a la educación superior, ya sean estas académicas, sociales o familiares, se revelan significativas para este análisis.

La construcción del modelo de investigación se sustentó en una exhaustiva revisión bibliográfica que permitió identificar variables explicativas de la deserción universitaria en el contexto nacional. Luego, a partir de la sistematización de antecedentes socioeducativos de los estudiantes de las cohortes 2014 y 2015 disponibles en los sistemas informáticos de la universidad, se consolidó una matriz tripartita con los datos asociados a las variables que emergieron del análisis de los referentes consultados.

Consecuentemente, se analizó la relación de cada una de las variables explicativas del estudio con la variable control deserción estudiantil. El análisis bivariado permitió identificar diecisiete variables, significativamente asociadas con la deserción estudiantil y precisar las relaciones de dependencia con el abandono de estudios.

El modelo multivariado predijo en un 86,4 % la conducta de abandono señalando siete variables categóricas independientes que, finalmente, se develan como factores relevantes del modelo predicción. Las variadas y sustentadas interpretaciones entregadas en los resultados del modelo, así como las sugerencias propuestas para mejorar el índice de retención universitaria, entregan un valor directo al estudio dirigido a optimizar uno de los indicadores más importantes vinculados a una gestión de calidad en las universidades, como es, la retención estudiantil.

La actual investigación aportó unidades teóricas por desiguales autores registrados sobre modelos predictivos que se reconcilian a la emprendedora deserción estudiantil una de las variables objeto de estudio, la cual se orienta al nivel de calidad en la ejecución de este modelo multivariado, igualmente integra los apoyos teóricos sobre los factores determinantes de la deserción como son académicos, sociales o familiares, para así manejarlos y equilibrarlos con las discusiones, conclusiones y recomendaciones de este estudio.

Finalmente, Rodríguez et al., (2017), titularon su monografía "Modelo predictivo para la permanencia en la educación superior Línea Temática: Factores asociados. Tipos y perfiles de abandono, en la Universidad de Playa Ancha de Ciencias de la Educación – Valparaíso, Chile. El objetivo de este trabajo es reportar el modelo multivariado predictivo para la permanencia universitaria a partir de atributos previos al ingreso a la universidad y de variables medidas durante el primer semestre de 2017 de los estudiantes de la Universidad de Playa Ancha que ingresaron a la universidad a través del Programa de Acompañamiento y Acceso Efectivo a la Educación Superior del Ministerio de Educación de Chile.

El presente texto describe de manera breve el modelo de Tinto (1975). Además, presenta un modelo adaptado por el PACEUPLA, la metodología de análisis de datos, y los resultados obtenidos tras un semestre de implementación. Se analizan características contextuales y de escolaridad previa a la universidad, el compromiso del estudiante con la universidad, variables del sistema académico, y la integración universitaria de los estudiantes.

El análisis de datos buscó identificar variables que tienen mayor influencia en la permanencia o deserción de los estudiantes, generando un modelo multivariado predictivo para la permanencia estudiantil. Se definió la variable Rendimiento Medio y una regresión lineal sobre ella. De los análisis se desprende que los predictores de efectos significativos son el promedio de notas de la enseñanza media (NEM), la asistencia a clases (ASISTENCIA) y la cantidad de asignaturas aprobadas (APROB), todas con una significancia del 5%. Se desarrolló un ajuste al modelo sólo considerando los predictores de efectos significativos.

El nuevo modelo obtuvo un coeficiente de determinación de 0.89, donde la variabilidad del Rendimiento Medio es explicada por los predictores NEM (0.08777), asistencia (0.42711) y aprob (-2.72077), todos significativos al 5%, sin sobreposición. Se definieron nuevas regresiones, una con un nivel de explicación

de 36% siendo significativo (p-valor =0.0004) y una tercera cuyo poder predictivo es del 75%, siendo altamente significativo (p-valor=0.000).

El modelo PACEUPLA dejó fuera factores que en otras investigaciones han sido fundamentales, como el Puntaje en la PSU, otorgando mayor impacto al NEM. Con esto, el modelo da mayor valor a la trayectoria escolar, dado que resulta más significativa en la predicción del rendimiento medio de los estudiantes en la educación superior.

Este artículo aportó propuestas sobre modelos predictivos, los cuales corresponden a una de las dimensiones de la presente investigación, la cual se describe a la técnica o modelo multivariado predictivo para la permanencia estudiantil, lo cual atribuye una muestra para la elaboración de las recomendaciones de la presente investigación.

2.1.2 Fundamentación teórica

2.1.2.1 Deserción estudiantil

Según el Ministerio de Educación Nacional República de Colombia (2019), señala que la deserción estudiantil es el abandono del sistema escolar de los estudiantes, provocado por la combinación de factores que se generan tanto en el interior del sistema como en contextos de tipo social, familiar, individual y del entorno. La tasa de deserción intra-anual solo tiene en cuenta a los alumnos que abandonan la escuela durante el año escolar, esta se complementa con la tasa de deserción interanual que calcula aquellos que desertan al terminar el año escolar.

Por otro lado, Ecured (2019), menciona que la deserción escolar es un término utilizado en Latinoamérica para referirse al abandono de la escuela. Se trata de aquella situación en la que el alumno después de un proceso administrativo de

separación o retiro, finalmente, comienza a retirarse antes de la edad establecida por el sistema educativo sin obtener un certificado.

Debido a los contextos sobre la deserción estudiantil se puede definir como el abandono de las actividades escolares por parte del estudiante afectado por factores ya sean individuales, académicos, institucionales o socioeconómicos o por el simple hecho de cambiar los estudios por negocios o deseos de tener un empleo.

2.1.2.1.1 Factores determinantes de la deserción

El Ministerio Nacional de educación (MEN). (2014), desarrolla un resumen del estado del arte de los factores determinantes de la deserción agrupando las variables más utilizadas en cuatro categorías: individuales, académicos, institucionales y socioeconómicos. Igualmente, afirma que existe consenso en que la deserción estudiantil es el resultado del efecto no de una sola categoría, sino del efecto individual y de la interacción de diferentes categorías de factores.

Sin embargo, en términos generales y de acuerdo con la revisión de la literatura, se puede decir que existen más trabajos que destacan la perspectiva institucional y en los que los diferentes conjuntos de variables (institucionales, socioeconómicas, académicas y personales) son analizados de manera independiente y no como un conjunto de factores que determinan la decisión de desertar. Debido a ello, Castaño et al., (2007), resumen las perspectivas del análisis en el estudio de la deserción; además, se agrupan las variables más utilizadas en las cuatro categorías relevantes o determinantes del problema.

Entre estas variables están: género, raza, discapacidad, localización de la vivienda, edad de entrada a la institución, colegio y puntaje en los exámenes de Estado. En este sentido, se ha encontrado que los factores más relevantes para explicar la deserción son: carencia de tiempo, escasa tutoría, poca información

sobre el proceso de enseñanza-aprendizaje, falta de soporte y dificultad de comunicación con las instituciones (Fozdar et al., 2006, 37).

Consecuentemente, Guzmán et al., (2009), puntea el objeto de identificar los determinantes de la deserción estudiantil para la educación superior de Colombia, el Sistema de Prevención de la Deserción en Educación Superior –SPADIES ha incluido variables individuales, socioeconómicas, académicas e institucionales, que se muestran a continuación, y las cuales son cuantificables y están disponibles en la información suministrada por las instituciones de educación superior, el ICFES y el Instituto Colombiano de Crédito y Estudios Técnicos en el Exterior ICETEX, así como por el Sistema Nacional de Información de la Educación Superior SNIES.

Estos factores son: Personales, académicas, socioeconómicas, e institucionales; dentro de los mismos se encuentran: la edad, sexo, tasa de repitencia, área del conocimiento, puntaje, origen, carácter, apoyo financiero, crédito, vivienda propia, nivel de ingresos, número de hermanos, posición dentro de los hermanos, educación de la madre, tasa de desempleo departamental, ubicación de la institución de educación superior, problema de índole netamente económico, relacionado con la imposibilidad del estudiante de garantizar su vinculación al sistema, dadas las dificultades y limitaciones para cubrir sus costos de matrícula y sostenimiento.

Sin embargo, la evidencia de un estudiante con bajas condiciones académicas, ingresos bajos del núcleo familiar (entre uno y dos salarios mínimos), compromisos laborales y económicos, madre de nivel educativo de básica primaria, poseer vivienda propia o en arriendo, estudiantes que no tienen hermanos, entre otros.

Por otro lado, se evidencia una variabilidad de factores individuales, académicos, socioeconómicos e institucionales por medio de diferentes autores:

a. Sexo: Predomina un mayor riesgo de deserción relativo para los varones, explicado tentativamente por una actitud menos perseverante de los varones. Lopera (2008), Castillo et al., (2010) y Giovanoli (2002) obtienen mayor riesgo de deserción en los varones, pero el coeficiente no resultó significativo.

b. Lugar de origen: Se considera una variable dicotómica que distingue a quienes provienen de la ciudad donde está emplazada la sede la Universidad del resto de los alumnos. Se espera tanto para los alumnos con elevados tiempos de traslado como para quienes necesitaron migrar, una potencial mayor propensión a abandonar. (Oquendo, 2008, p. 35); (Castillo et al., 2010, p. 39).

c. Grado de responsabilidad: Esta variable intenta captar rasgos de personalidad (comportamientos estables y duraderos en el individuo a lo largo del tiempo), que implican coherencia, organización, perseverancia y posibilidades de sostener el esfuerzo en la actividad académica. Se mide a través de reactivos autoevaluados. (Lopera 2008, p. 24); (Castillo et al., 2010, p. 42).

d. Estrato social (nivel socioeconómico): En trabajos sobre deserción (Ishitani, 2006; Castillo, 2010) han detectado un menor riesgo de abandono en estudiantes provenientes de hogares de mayor estrato socioeconómico (padres más escolarizados y/o con mejor inserción laboral).

e. Situación laboral del estudiante: El hecho de destinar horas a la actividad laboral aumenta significativamente el riesgo de abandono (Giovanoli, 2002; Castaño et al., 2010). En este estudio identifica si el estudiante es empleado en relación de dependencia o no, e intenta captar la situación laboral que más restricciones puede imponer a un estudiante en cuanto a sus tiempos de estudios.

f. Nivel educativo del padre: Esta variable capta distintos elementos relacionados con la comprensión familiar de las responsabilidades y actividades propias de la vida universitaria. Quienes tienen padres con estudios universitarios en general

asignan importancia a la culminación de estudios, así como brindan el apoyo necesario para el desarrollo de actividades como estudio en la casa, reuniones de trabajo en grupo, etc. (Giovanoli, 2002, p. 27).

g. Residir solo: Giovanoli (2002), encuentra evidencia que indica que los estudiantes que viven solos tienen mayor riesgo de desertar que quienes residen con sus familias.

h. Situación residencial: Las características habitacionales favorables se espera que influyeran en sentido positivo la permanencia en la universidad. (Giovanoli 2002, p. 29).

i. Tener experiencia universitaria previa: el haber cursado previamente alguna otra carrera o la misma carrera en otra institución se relaciona en otros estudios empíricos con una mayor propensión a la deserción. (Giovanoli 2002, p. 33).

j. Resultado en el DAT – Razonamiento verbal: La relación que se espera detectar es un menor riesgo entre los estudiantes que obtuvieron mayor puntaje en el DAT. (Giovanoli, 2002, p. 37).

k. Rendimiento académico en el cuatrimestre (consta de dos aspectos, el referido al número de materias regularizadas y el promedio de notas obtenidas en los parciales de las materias). El rendimiento académico corresponde a los logros demostrados en términos de resultados del proceso de aprendizaje en cuanto a los contenidos de las diferentes materias. (Castaño et al., 2010, p. 54).

l. Seguridad en la elección de carrera: Se espera que la proporción de abandonos sea superior entre quienes manifiestan algún grado de inseguridad sobre la elección vocacional. Esta relación ha sido ampliamente documentada en estudios vinculados con la orientación vocacional (Boado, 2005, p. 23; Cabrera et al., 2006, p. 55).

m. Bienestar psicológico: Este indicador construido a partir de las auto-percepciones correspondientes a reactivos vinculados con el manejo de estrés, autoestima, predisposición al optimismo / pesimismo, etc. (Boado, 2005, p. 28); (Cabrera et al., 2006. p. 35).

n. Integración: El indicador sintetiza la propia visión sobre la manera de relacionarse con pares, las preferencias por compartir actividades con otros, etc. (Boado 2005, p. 31); (Cabrera et a., 2006, p. 39).

o. Apoyo familiar: Grado en que percibe que la familia estimula y apoya para que desarrolle los estudios. (Boado 2005, p, 32); (Cabrera et al., 2006, p. 40).

p. Sentimiento de añoranza: Grado de intención de retornar a su lugar de origen. (Boado 2005, p. 32); (Cabrera et al., 2006, p. 40).

Estos factores no serán utilizados dentro del desarrollo de la investigación, pero las categorías o variables que se utilizarán serán las que son descritas por la diversidad de autores, y no las disponibles en el Sistema de información Académico (SIA) que dispone en la Sede Maicao en los resultados recopilados de los procesos de Admisión a la Universidad de la Guajira y en los datos extraídos, las cuáles son muy pocas, debido a ello, estos factores serán descritos más adelante. A continuación, se describen los diferentes factores determinantes:

A. Factores Individuales

En algunos estudios se menciona que el abandono de la educación superior obedece a dificultades, exigencias que se les impone al estudiante, lo cual da lugar a la explicación de la deserción a partir de factores individuales y contextuales que infieren en la conducta del estudiante (Gil et al., 2001, p. 38).

Autores como Spady (1970), Tinto (1975), Ethington, (1970), Giovagnoli, (2002) menciona al respecto factores como la personalidad, dificultades de adaptación personal a la Institución, incompatibilidad entre la vida académica y las exigencias del trabajo, desencanto entre la vida académica y las exigencias del trabajo, dificultad en la relación enseñanza-aprendizaje, añadiendo otros factores como la edad, género y estado civil, posición entre los hermanos, entorno familiar, calamidad y problemas de salud, integración social.

En lo que a la familia respecta, se reconocen las presiones que los padres ejercen en la selección vocacional de sus hijos y la ausencia de apoyo en los procesos de formación, bien sea porque los miembros del grupo familiar no se involucran o por falta de herramientas para acompañar el proceso. A esto debe añadirse circunstancias como cambios de trabajo, divorcios, embarazo, etc., asuntos que causan efectos emocionales y económicos en la vida de los estudiantes y de sus familiares.

Dado a la variabilidad de factores individuales desarrollados por varios autores, se determina la pauta por Spady (1970), Tinto (1975), Ethington, (1970), Giovagnoli, (2002), los cuales han confirmado que los mismos determinan un alto porcentaje la deserción estudiantil, lo cual muestra controversias a lo observado en el Sistema de Información Académico (SIA) los resultados recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao.

B. Factores Académicos

Desde la orilla de lo académico emerge otro factor decisivo en la deserción estudiantil. Se alude tanto a los vacíos en conocimiento, habilidades cognitivas, problemas de aprendizaje, deficiencias en los niveles de lectura inferencial y crítica, como de escritura y desarrollo del pensamiento lógico-matemático, como a la ausencia de hábitos de estudio, tipo de colegio, calidad del programa, método de estudio, número de materia, insatisfacción con el programa, disciplina,

entusiasmo, compromiso, exigencia, perseverancia, motivación para aprender, entre otros. Y, yendo aún más allá, a la carencia de estilos de vida que propulsen el éxito académico. (Báez y Pedraza, 2011, p. 48).

De igual manera dentro del conjunto de variables académicas, las que parecen ser importantes en la explicación del riesgo de deserción son: el desempeño académico en la universidad. Los resultados muestran que, a mayor número de créditos cursados, menor es el riesgo de desertar, mientras a mayor número de créditos reprobados, es decir a mayor repitencia, mayor es el riesgo.

Así mismo, el grado de satisfacción del estudiante con el programa parece ser una variable que incide directamente en el riesgo de desertar, al encontrarse que aquellos estudiantes con un nivel de satisfacción alto o medio tienen menor riesgo de desertar en comparación con quienes tienen un bajo grado de satisfacción (Castaño et al., 2007, p. 57).

En relación a los factores académicos desplegados por varios autores, se establece la señalada por (Báez y Pedraza, 2011), los cuales han permitido que los mismos determinan un alto porcentaje en la determinación de la deserción estudiantil, lo cual pauta polémicas a lo observado en el Sistema de Información Académico (SIA) resultados recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao, evidenciando que el mismo posee pocos factores académicos recolectados.

C. Factores Institucionales

Yorke y Longden (2004), mencionan que en este campo de factores institucionales se destacan la calidad de los programas y los procesos de formación, la flexibilidad curricular, las prácticas pedagógicas de los docentes, becas y formas de financiamiento, orden público, entorno político, nivel de interacción personal,

apoyo académico y psicológico y todo aquello que estimule la vinculación del estudiante con el medio académico e investigativo.

La calidad de la experiencia educativa incluye elementos como el tipo de enseñanza, la organización del programa académico, el apoyo de los miembros de la comunidad fuera del aula de clase y el número de estudiantes por clase, y que innegablemente estos ejercen gran influencia en la motivación de los estudiantes por continuar vinculados con sus opciones de formación como: normalidad académica, becas y formas de financiamiento, recursos universitarios, orden público, entorno político, nivel de interacción personal con profesores y estudiantes, apoyo académico y apoyo psicológico (Adelman 1999, p. 23).

Debido a los factores institucionales desarrollados por varios autores, se establece la señalada por (Báez y Pedraza, 2011), los cuales han reconocido que los mismos comprueban un alto porcentaje en la determinación de la deserción estudiantil, lo cual modela cuestiones a lo visto en el Sistema de Información Académico (SIA) recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao, demostrando que el mismo tiene escasos factores institucionales en el Programa de Ingeniería en Sistemas.

D. Factores Socioeconómicos

Los factores socioeconómicos se encuentran: El Estrato, situación laboral, situación laboral de los padres e ingresos, dependencia económica, personas a cargo, nivel educativo de los padres y entorno macroeconómico. (Tinto 1995, p. 44; Gaviria 2002, p. 33). Estos factores presentan consecuencias como: dificultades de financiamiento, dificultades con el crédito, falta de trabajo, la falta de financiamiento en universidades que no tiene programas de ayudas becarias reembolsables y si existen tienen una alta limitación por el capital que manejan estos programas.

Asimismo, se observa la alta tasa de desempleo que existe en el país, las expectativas de graduarse de la universidad y obtener un trabajo que le permita tener un nivel de vida adecuado es bastante desalentador por las desigualdades que existen. El inicio o formación de una nueva familia, es una limitante para seguir sus estudios exitosamente, debido a los nuevos compromisos y responsabilidades (Abensur 2009, p. 45).

Adicionalmente, Jiminian (2006), muestra las consecuencias más relevantes de los factores económicos, dadas que las mismas ocurren en el momento de ingresar a la educación superior, pues en este período la mayoría de los alumnos deben tener en cuenta su situación económica para estructurar sus decisiones. La poca capacidad adquisitiva y de solvencia frente a los costos educativos para los profesionales en formación se convierten en situaciones de constantemente evidenciado en lo que concierne a la influencia de dichos factores.

El asunto financiero es determinante en la posibilidad de vincularse y mantenerse en el sistema educativo. Si las instituciones y el sistema en general no cuentan con políticas claras de financiación y con las sinergias de interacción necesarias entre las diferentes dependencias será muy difícil de lograr el acceso y la permanencia de los estudiantes. (Báez y Pedraza 2011, p. 55).

Acceder, mantenerse y graduarse requiere inversión básicamente en matrícula, transporte, libros y alimentación. Los hogares de estratos 1 y 2 deben comprometer su patrimonio para tal efecto; es fundamental la estabilidad de los ingresos, si estos caen entonces los rubros educativos disminuyen para mantener otras necesidades básicas (Báez y Pedraza 2011, p. 55).

Planteado los factores socioeconómicos por varios autores, se instituye la descrita por Tinto (1995) y Gaviria (2002), los cuales han reconocido que los mismos patentizan un resultado significativo en la visualización de la deserción estudiantil en la educación superior, lo cual forma diferencias a lo observado en el Sistema

de Información Académico (SIA) recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao, manifestando que el mismo tiene insuficientes respuestas de los factores socioeconómicos en el Programa de Ingeniería en Sistemas.

2.1.2.2 Minería de datos

La minería de datos se refiere a la extracción "o la minería" del conocimiento de los grandes volúmenes de datos. El término es en realidad un nombre poco apropiado, la minería de datos de una manera más apropiada debería haber sido llamada "la minería de conocimiento de datos", que es lamentablemente largo. "La minería de conocimiento", un término más corto, no puede reflejar el énfasis en la minería de los volúmenes de datos. (Han et., 2011, p. 8).

Asimismo, Han et., (2011, p. 11), marca que muchas personas tratan la minería de datos como un sinónimo de otro término utilizado popularmente, el descubrimiento de conocimiento de datos, o KDD, consiste en una secuencia iterativa de los siguientes pasos: 1. Datos de limpieza, integración de datos, la selección de datos, transformación de datos, la minería de datos, modelo de evaluación y la presentación del conocimiento, el cual es considerado como un modelo predictivo de minería de datos para estimar la deserción estudiantil.

Por otro lado, Morales (2009), señala que la Minería de Datos es un campo en pleno desarrollo en el que se aplican métodos de varias disciplinas como los presentes en sistemas de bases de datos, data warehousing, estadística, el aprendizaje automático, visualización de datos, obtención de información y computación de alta performance.

Además, se utilizan métodos de las áreas de redes neuronales, reconocimiento de patrones, análisis espacial de datos, bases de datos de imágenes, procesamiento de señales y programación lógica inductiva; en tal sentido, numerosos

especialistas señalan que la Minería de Datos necesita de la integración de enfoques de múltiples disciplinas. Una de sus características principales es que invierte la dinámica del método científico. Es decir, primero se coleccionan los datos y luego se los “escucha” para que de ellos emerjan las hipótesis. Luego se validan esas hipótesis en los datos mismos.

Además, Valero (2009), plantea que la Minería de datos es una subdisciplina de las ciencias de la computación que presta apoyo a otras ciencias; su mayor fortaleza radica en la búsqueda de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles para la toma de decisiones en cualquier sector.

Es por ello que a través de la Minería de Datos, armada en procedimientos matemáticos de razonamiento, es viable llevar a cabo el análisis de amplios volúmenes de información recopilada en las bases de datos de las instituciones académicas; y manifestar modelos, agrupaciones, cambios o incoherencias, sobre datos que tratados de forma encerrada, no contribuyen componentes que admitan tomar decisiones para investigar las razones de deserción estudiantil; procedidas por ejemplo del vacío estudio de los expedientes académicos del alumnado. Caso que no desarrollado en el Programa de Ingeniería en Sistemas de la Universidad de la Guajira Sede Maicao.

2.1.2.3 Técnicas de Minería de Datos

En esta fase se realiza el modelamiento propiamente tal a través de la aplicación de distintas técnicas. El objetivo es extraer patrones y conocimientos previamente desconocidos, los cuales se obtendrían a través del procesamiento de los datos. Las técnicas aplicadas pueden ser de caracterización, asociación, clusterización, clasificación y regresión (Han y otros 2011, p. 44).

Además, señala que las tareas relacionadas con la clasificación hacen referencia a la construcción de modelos donde se categorizan los registros en grupos predefinidos o clases ya conocidas. En general, la variable dependiente en los modelos de categorización es del tipo categórico o discreto. Por otro lado, cuando se habla de actividades de regresión, se hace referencia al uso de modelos para predecir variables del tipo real. La decisión de aplicar tareas de categorización o regresión dependen crucialmente del objetivo que se busca.

Seguidamente, Han y otros, (2011), muestran que en la actualidad existen variadas técnicas de minería de datos que permiten la ejecución de las tareas anteriormente descritas. Respecto de la categorización y regresión es posible identificar máquinas de aprendizaje que, gracias a los avances tecnológicos, la implementación de estas se hace cada vez más accesible para los usuarios. Máquinas de aprendizaje tales como *Support Vector Machine*, *Decision Tree* y *Artificial Neural Net* han sido programadas y utilizadas como librerías en *softwares* de minería de datos.

Las técnicas que conforman el campo de la Minería de Datos buscan descubrir, en forma automática, el conocimiento contenido en la información almacenada en las bases de datos de las organizaciones. Por medio del análisis de datos, se pretende descubrir patrones, perfiles y tendencias. Es importante que estas técnicas sean las adecuadas al problema abordado. En este sentido, se pueden establecer dos grandes grupos de técnicas o métodos analíticos: los métodos simbólicos y los métodos estadísticos. Entre los métodos simbólicos se incluyen a las redes neuronales, algoritmos genéticos, reglas de asociación, lógica difusa, entre otros.

1. Técnica de Árboles de decisión

Un árbol de decisión es un diagrama de flujo, con estructura de árbol, en donde los nodos internos representan validaciones sobre los atributos, las ramas

representan las salidas de las validaciones, y los “nodos hoja” representan las clases. El nodo en la parte superior del árbol se le conoce como nodo raíz. Para clasificar una instancia “desconocida”, se sigue el flujo del árbol hacia abajo, de acuerdo a los valores que tengan los atributos para cada nodo, y cuando se llega a un “nodo hoja”, la instancia se clasifica de acuerdo a la clase asignada por dicho nodo. (Heredía y Nieto 2011, p. 47).

Existen diversos métodos para la inducción de árboles de decisión (ID3, C4, C4.5, Bayesiano, CART, etc.), cada uno de ellos ofrece diferentes capacidades, pero en general, dichos algoritmos son apropiados para solucionar «problemas de clasificación». El algoritmo ID3. La idea básica del algoritmo ID3 tiene su fundamento en la iteración. Un subconjunto del conjunto total de datos de entrenamiento, al cual se le conoce como Windows, es elegido de manera aleatoria para formar un árbol de decisión; este árbol clasifica de manera correcta todos los objetos que pertenecen a Windows.

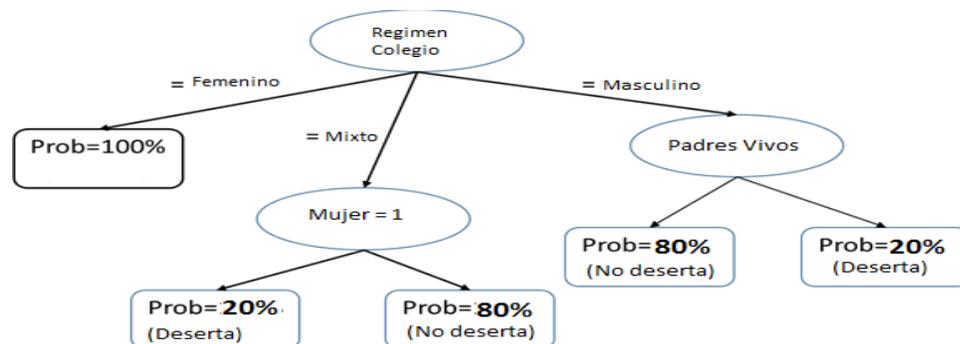
El resto de los objetos, dentro del conjunto de datos de entrenamiento, es clasificado utilizando dicho árbol. Si el árbol da una respuesta correcta para todos estos objetos, entonces también es correcto para el conjunto total de datos de entrenamiento, terminando así el proceso. Si no, una selección de objetos clasificados incorrectamente es adicionada al subconjunto window, y el proceso continúa.

El algoritmo C4.5. El algoritmo forma parte de la familia de los «Árboles Inducidos de Arriba hacia Abajo» (Top Down Induction of Decision Trees, TDIDT por sus siglas en inglés). Pertenece a los métodos inductivos del Machine Learning, los cuales aprenden a partir de ejemplos preclasificados. Tanto el algoritmo ID3 como el C4.5 fueron propuestos por Ross Quinlan. El algoritmo C4.5 es una extensión del algoritmo ID3, el cual trabaja únicamente con valores discretos en los atributos. En cambio, el algoritmo C4.5 permite trabajar con valores continuos, separando los posibles resultados en dos ramas: una para aquellos y otra para todos los de

esta forma, C4.5 genera un árbol de decisión a partir de datos mediante particiones realizadas de manera recursiva.

Asimismo, los árboles de decisión, o *Decision Tree* (DT) en inglés, en la disciplina de minería de datos es considerado un modelo de clasificación planteado por (Quinlan, 1986), el cual se basa en las teorías de decisiones para realizar clasificaciones a las base de datos en donde se apliquen algoritmos minería de datos. Quinlan ha realizado grandes aportes a los algoritmos de árboles de decisiones, siendo los más reconocidos el C4.5 e ID3.

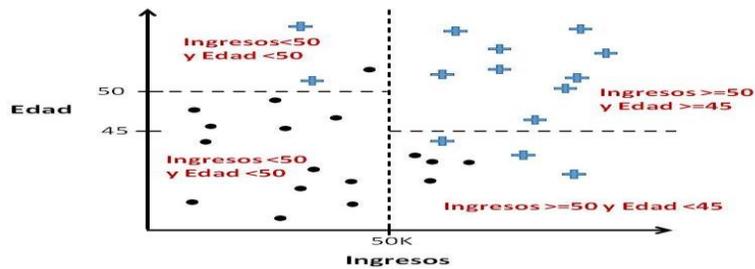
Los árboles de decisión segmentan los registros utilizando técnicas matemáticas y estadísticas, introduciendo el concepto de entropía (índice de incertidumbre o desorden), el cual sirve para identificar el siguiente atributo de segmentación. Gráficamente los árboles se componen de nodos, ramas y hojas. Los nodos son puntos de unión, en donde se refleja una toma de decisión. En el libro *Data Science for Business* (Provost y Fawcett 2013, p. 32), se muestran dos claros ejemplos de la segmentación y representación gráfica de un árbol de decisión. Respecto de la Figura 1, se puede ver la secuencia de nodos que reflejan la decisión de segmentación de un conjunto de datos. Los nodos finales reflejan la clasificación final con la probabilidad correspondiente.



Fuente: Provost & Fawcett, (2013).

Figura 1. Representación gráfica del resultado obtenido por un algoritmo basado en la máquina de aprendizaje Árbol de Decisiones.

Además, en la Figura 2 se muestra de manera gráfica la segmentación realizada a través del árbol de decisión. Cada división refleja un nodo y los puntos dentro de un cuadro reflejan las hojas o nodos finales del árbol.



Fuente: (Provost & Fawcett, 2013)

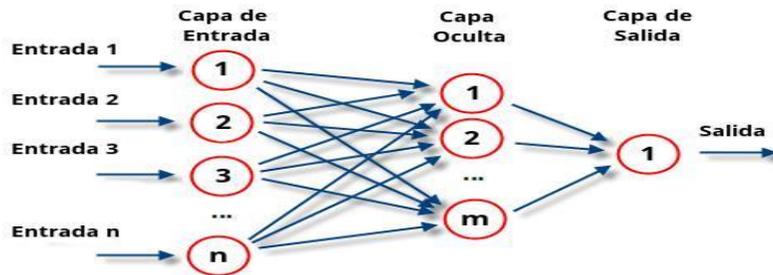
Figura 2. Representación gráfica de la segmentación según algoritmo de Árbol de decisión

2. Técnica de Redes Neuronales Artificiales

Las redes neuronales artificiales o artificial neural net (ANN) fueron introducidas inicialmente como concepto de red neuronal por los neurólogos (McCulloch & Pitts, 1943). Quince años más tarde, (Rosenblatt, 1958) generó el primer perceptrón simple basado en los conceptos de red neuronal, proponiendo así los fundamentos de una red neuronal artificial. Una red neuronal se compone de redes de nodos llamados neuronas. Cada nodo recibe un conjunto de entradas provenientes de otros nodos y entregan una salida. Esta salida se compone por tres funciones. (Kumar y Otros 2012, p. 31):

- 1. Función de propagación:** Es la función que se compone por la sumatoria de las entradas multiplicados por un peso de interconexión.
- 2. Función de activación:** Es la función que modifica la función anterior (aprendizaje). Puede que la configuración de la red no tenga esta función, por lo que la salida es la misma función de propagación.
- 3. Función de transferencia:** Es la función que se aplica al valor que entrega la función de activación. Su principal es para acotar el rango de salida del nodo. Las más comunes son la función sigmoidea (intervalos entre 0 y 1) y la tangente

hiperbólica (intervalos entre -1 y 1). A modo de ejemplo, la Figura 3 muestra la estructura básica de una red neuronal.



Fuente: Kumar y Otros (2012).

Figura 3. Ejemplo de estructura de una red neuronal artificial

3. Técnica de Regresión Logística

La Regresión Logística o *Logistic Regression* (LR) en inglés, es un caso especial de regresiones cuyo uso es para predecir el resultado de una variable dependiente categórica. Tiene bastante uso en los cálculos de probabilidades, donde se predice la ocurrencia de un evento en función de otros factores. A modo de ejemplo, supongamos que la variable de respuesta y toma valores 0 y 1. (Kumar y Otros 2012, p. 32). De acuerdo a lo que postula la regresión logística, la probabilidad posterior $P(y|x)$ de respuesta a la variable condicionada del vector x sigue una función logística definida como:

$$P(y = 1|x) = \frac{e^{w'x}}{1 + e^{w'x}}$$

$$P(y = 0|x) = \frac{1}{1 + e^{w'x}}$$

En las ecuaciones previamente mostradas, el algoritmo identifica los coeficientes w de forma iterativa, usualmente a través del método de máxima verosimilitud. En

general los modelos y algoritmos de regresión logística presentan la misma dificultad que las regresiones lineales, en otras palabras, pueden adolecer de problemas de multicolinealidad y sesgo.

Parámetros

En la etapa de entrenamiento, los algoritmos de regresiones logísticas obtienen las predicciones para un conjunto de observaciones X y al igual que los *support vector machines*, es posible calcular el costo de error de las clasificaciones erróneas. Adicionalmente, los algoritmos iteran con el objetivo de identificar el mejor conjunto de los valores w_i , en donde el costo total de clasificaciones erróneas busca ser minimizado.

Asimismo, el análisis de regresión se utiliza para modelar relaciones entre variables y para realizar pronósticos o predicciones de respuestas a partir de variables explicativas y de este modo crear un modelo donde se seleccionan las variables que están vinculadas con la respuesta, descartando aquellas que no aportan información.

Conjuntamente, permite detectar interacciones entre las variables independientes que afectan a la variable dependiente o predicha. La regresión logística es un tipo especial de regresión que se utiliza para explicar una variable dicotómica en función de variables independientes que pueden ser cuantitativas o cualitativas. (Castejón 2011, p. 22). El modelo de regresión logística establece la relación entre la probabilidad de que ocurra el suceso dado que el individuo presenta los valores $(X = x_1, X = x_2, \dots, X = x_k)$:

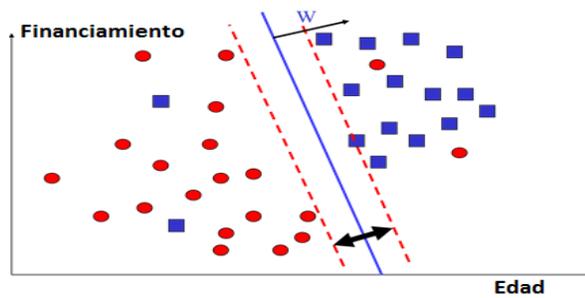
$$P[Y = 1 / x_1, x_2, \dots, x_k] = \frac{1}{1 + e^{(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_k x_k)}}$$

4. Técnica Support Vector Machine

El Support Vector Machine (SVM) fue introducido por (Vapnik & Chervonenkis, 1964). Hoy en día, desde la disciplina de máquinas de aprendizaje el SVM es un modelo del tipo de aprendizaje supervisado el cual utiliza algoritmos de clasificación y análisis de regresión. En términos prácticos, se puede explicar el funcionamiento del SVM como la clasificación de un conjunto de registros a través de la separación con un hiperplano, el cual minimiza el costo de error de clasificación de cada registro a una de las dos clases en estudio. (Vercellis 2009, p. 20).

Asimismo, el autor Vercellis (2009), muestra que formalmente se define el Support Vector Machine como la función de separación representada por un hiperplano en el espacio R^n , el cual maximiza el margen de separación. En una separación de puntos por dos clases, el problema se denomina del tipo linealmente separable. Bajo este concepto, en un espacio R^n con finitos puntos que representan observaciones, estos pueden ser separados por infinitos hiperplanos. Sin embargo, los algoritmos que aplican Support Vector Machine identifican la separación lineal óptima, el cual está definido como la máxima capacidad de generalización y el mínimo error empírico de clasificación.

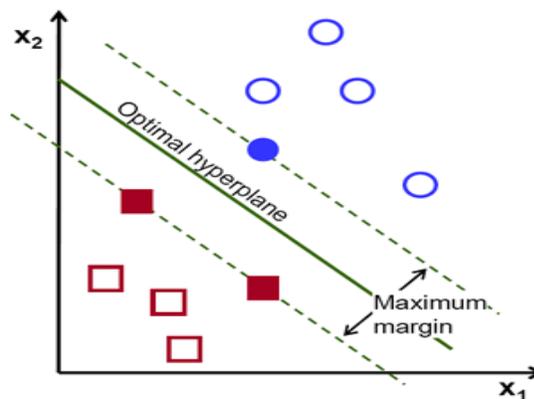
El margen de separación es definido como la distancia entre el par de paralelos canónicos, por lo que el margen es igual a dos veces el mínimo entre los puntos de entrenamiento y el hiperplano de separación. Estos puntos los cuales se minimiza la distancia de separación del hiperplano, son llamados vectores de soporte (*support vectors*). Sobre estos se obtienen las reglas de clasificación (Vercellis, 2009). Matemáticamente, si se define w como el vector de coeficientes del hiperplano y b como el intercepto, se define el hiperplano dado como: $w'x = b$. Mientras que los dos hiperplanos paralelos canónicos son: $w'x - b - 1 = 0$. $w'x - b + 1 = 0$.



Fuente: Vercellis, (2009).

Figura 4. Representación gráfica de la aplicación del algoritmo SVM

Los parámetros del Support Vector Machine son los costos asociados al error de clasificación. Por lo tanto, en esta tesis se optimizará el C que otorgue el mejor desempeño de los modelos. La SVM se basan en los fundamentos de la teoría de aprendizaje estadístico desarrollada por Vapnik y pertenecen a la categoría de los clasificadores lineales ya que establecen separadores lineales o hiperplanos, ya sean en el espacio original de los datos de entrada, o en un espacio transformado. (Carmona 2014, p. 11). Dado un conjunto de vectores x_1, \dots, x_n se define separable si existe algún hiperplano en $X \times X$ con etiqueta $\{+, -\}$ $1 \leq i \leq n$ y y_i de aquellos con etiqueta $i \in \{+, -\}$ $x_i \cdot y_i > 0$ donde $1 \leq i \leq n$ $y_i \in \mathbb{R}$ que separa los vectores $x_i \in \mathbb{R}^d$ $y_i \in \mathbb{R}$ para $y_i \in \{+, -\}$.



Fuente: OpenCV dev team (2014).

Figura 5. Hiperplano separador en R2 SVM

La ecuación del Hiperplano está dada por $(x - x_0)' \cdot \frac{1}{\|b\|} b = b$ donde $d \in \{1, -1\}$ y $\{x_0, b\} \in \mathbb{R}^n$, es un número real. Si los datos no son separables linealmente en el espacio se pueden utilizar funciones Kernel (no lineales).

- Polinomial $k(x, x') = \sum_{i=1}^n x_i x'_i + c$ (2-2)
- Radial Gaussiana $k(x, x') = e^{-\gamma \|x - x'\|^2}$
- $(\cdot)^2$ $k(x, x') = \sum_{i=1}^n x_i x'_i + c$ (2-3).

Estas funciones resuelven el problema de clasificación trasladando los datos a un espacio donde el hiperplano de solución es lineal y, por tanto, más sencillo de obtener. Una vez conseguido, la solución se transforma, de nuevo, al espacio original. El SVM se hace inicialmente con datos de entrenamiento para los cuales ya se conoce el valor de la variable dependiente, luego de este entrenamiento, se define un modelo que permitirá clasificar otro individuo. (Carmona 2014, p. 13).

2.1.2.4 Recursos Necesarios

2.1.2.4.1 Recursos Tecnológicos

Según la Universidad de Murcia (2014), muestra que los siguientes recursos tecnológicos para la minar datos o almacenamiento son el hardware y el software, los cuales se describen a continuación:

A. Hardware

El mismo se considera como:

- 1.- Ordenador de sobremesa. Ordenador montado por piezas (clónico) con arquitectura.
- 2.- Ordenador portátil (notebook). Ordenador de marca prestigiosa con arquitectura. Gama profesional.
- 3.- Tableta (tablet). Tableta de marca conocida y nivel medio de prestaciones. Sistema operativo Android.

4.- Teléfono móvil celular tipo Smartphone, con acceso Wifi y GPS. Teclado alfanumérico completo. De marca prestigiosa.

5.- Scanner portable. De marca prestigiosa. Admite hojas sueltas.

B. Software

En cualquier equipo informático es imprescindible una parte lógica, software, y dentro de este integrante se puede distinguir entre el sistema operativo, herramientas y lenguajes para el desarrollo, y aplicaciones para el usuario final. Por ejemplo, para un ordenador personal se corresponderían respectivamente con: Windows 8, entorno de desarrollo Eclipse y editor de textos Libreoffice Writer.

Por otro lado, la Enciclopedia de Ejemplos (2017), muestra el Hardware y Software de la siguiente manera: En computación, los términos hardware y software aluden a aspectos distintos de todo sistema informático: el aspecto físico y el digital respectivamente, cuerpo y alma de toda computadora. El hardware es el conjunto de las partes físicas que integran el cuerpo de un sistema computarizado: las placas, circuitos, mecanismos y dispositivos eléctricos, así como de procesamiento, soporte y conexión.

2.1.2.4.2 Recursos Humanos

Etxebarria (2017), menciona que los recursos humanos, las personas, son el principal capital innovador de una organización, asimismo, los recursos humanos de una organización son los capitales de energía física e inteligente, de los que son propietarios cada una de las personas que forman parte de la organización y que aportan en mayor o menor cantidad y calidad a la misma.

Del mismo modo, Chiavenato, (2009), define la gestión del talento humano como: el conjunto de políticas y prácticas necesarias para dirigir los aspectos de los

cargos gerenciales relacionados con las personas o recursos, incluidos reclutamiento, selección, capacitación, recompensas y evaluación de desempeño.

Por otro lado, Lledo, (2011), afirma que los recursos humanos tienen un enfoque de aplicación y practica de las actividades más importantes dentro de la organización o empresas siendo la gestión del talento humano un pilar fundamental para el desarrollo exitoso de los procesos, pues al final las personas son los responsables de ejecutarlas actividades porque los proyectos no se desarrollar por si solos. Dentro de los recursos humanos que se utilizan en la minería de datos se encuentran los ingenieros en sistemas, los técnicos en informática y analistas de datos, los cuales se describen a continuación:

A. Ingeniería en Sistema

Pérez y Merino (2009), mencionan que la ingeniería de sistemas es una carrera universitaria que se encarga del diseño, la programación, la implantación y el mantenimiento de sistemas. A diferencia de otras ramas de la ingeniería, esta disciplina no se ocupa de productos tangibles (los ingenieros civiles, por ejemplo, construyen edificios), sino de productos lógicos.

Seguidamente, lingen.unam (2018), señala que la ingeniería de sistemas es un campo de la ingeniería que se encarga del diseño, la programación, la implantación y el mantenimiento de sistemas. Utiliza un enfoque interdisciplinario que permite estudiar y comprender la realidad, con el propósito de implementar u optimizar sistemas complejos. Además, Emagister (2017), muestra que la ingeniería de sistemas es una rama de la ingeniería que se encarga del diseño, la programación, la implantación y el mantenimiento de sistemas. A diferencia de otras especialidades, esta disciplina no se ocupa de productos tangibles (los ingenieros civiles, por ejemplo, construyen edificios), sino de productos lógicos.

B. Técnicos electrónicos

Amador (2018), enfoca que los técnicos informáticos de servicios de soporte trabajan en colaboración con ingenieros informáticos y ayudan a los clientes o compañeros de trabajo a hacer funciona equipamientos informáticos. Deben instalar, probar, mantener y actualizar los equipamientos y software informáticos. Algunos técnicos informáticos de servicios de soporte proporcionan soporte técnico a través de líneas de atención telefónica.

Adicionalmente, Educaweb (2018), menciona que el Técnico en Informática es un programa diseñado para brindar una sólida formación profesional teórica y práctica en el manejo de las TIC's en ambientes de empresas públicas, privadas, mixtas o de autogestión, satisfaciendo así una demanda del mercado que requiere de personal académicamente formado, en las artes del software aplicado y de las tecnologías de la información aplicadas al mundo de las empresas. Asimismo, Fisc (2018), afirma que la Informática es la rama de la Ingeniería que estudia el hardware, las redes de datos y el software necesarios para tratar información de forma automática.

C. Analista de datos

Según Universia (2018), es una de las profesiones más demandadas por las empresas de la actualidad. La carrera de analista de datos tiene un papel fundamental en el desarrollo de las empresas. Los estudiantes que deseen formarse en una carrera de gran empleabilidad tienen en esta una excelente opción. El dominio del análisis de datos facilita el empleo en prácticamente cualquier empresa.

Consecuentemente, MBA & Educación Ejecutiva (2018), pauta que el analista de datos se ha convertido en un pilar fundamental para el crecimiento y desarrollo de una empresa. Toda la información relevante sobre clientes, mercado, rendimiento

de la empresa, de las inversiones publicitarias debe ser medida, procesada y analizada con el objetivo de tomar decisiones informadas que mejoren el rendimiento del negocio.

Igualmente, Emagister (2017), muestra que el data scientist o analista de datos como el experto en la gestión de la información compleja. Su objetivo es sonsacar datos relevantes para que las empresas puedan predecir algunos comportamientos de los usuarios. Por ejemplo, a qué hora van a consumir con mayor intensidad y qué productos serán los más demandados.

2.1.2.6 Modelos Predictivos

Según Guzmán et al. (2009), la deserción se encuentra la aplicación de diversas metodologías estadísticas con el fin de explicar el fenómeno. Los modelos de variable dependiente discreta o de respuesta cualitativa, y en particular, los modelos de regresión logit, probit y análisis discriminante han sido los más empleados para estimar el riesgo de desertar en un punto determinado del tiempo.

1. Modelos de retención

A lo largo de los años y relevando la problemática, varios autores contribuyeron en desarrollar hipótesis y estudios para determinar el modelo más adecuado frente a la retención estudiantil en la educación universitaria. El modelo que tal vez haya tenido la mayor aceptación, es el planteado por Vincent Tinto, que ha vinculado variadas teorías y de diferentes exponentes, y propone que el éxito de un estudiante está mediado por su grado de integración académica y social; es decir, por el cumplimiento de estándares académicos y por los lazos que se tejan entre él, su entorno y los agentes de la institución. En el año 2012, Tinto propone un cambio de enfoque sobre el fenómeno de persistencia y un modelo de acción institucional.

En este contexto, Tinto provee una distinción conceptual entre los términos de retención y persistencia. La persistencia la define desde la perspectiva del estudiante, quien tiene el control de permanecer o no en la institución y en la carrera elegida hasta graduarse. El autor reflexiona en torno a diversas condiciones mínimas que deben considerar las universidades para crear estas comunidades, entre las cuales destaca el esfuerzo sistemático de esclarecer lo que se necesita para tener éxito en el proceso de formación.

La idea esencial es proveer al estudiante información de calidad que le ayude a razonar por los diferentes cursos, a lo largo de la carrera y en su estancia en la universidad. Ya no interesa principalmente saber por qué el estudiante se va, sino saber qué estrategias pueden implementar las instituciones para que el estudiante se desarrolle cognitivamente y afectivamente en su progreso académico hasta la obtención del título universitario. (Cabrera et al., 2012, p. 27).

Entendiendo que la deserción de los estudiantes del sistema de educación superior es un problema que afecta a todos y es sinónimo de exclusión social, se torna urgente buscar las soluciones que aplaquen de mejor manera las causas de este fenómeno que afecta, no solo a la promoción de las personas, sino también, al desarrollo general de la sociedad.

Para superar el negativo diagnóstico que arrojan los actuales índices de deserción, es necesario indagar más allá de las cifras y poner énfasis en el conocimiento del perfil de los estudiantes, de sus características generacionales, económicas, sociales y académicas. Considerando todas estas variables, se podrá tener una visión más completa y significativa a la hora de tomar decisiones en la construcción de un modelo de retención que, sin duda, debe ser basado en programas de apoyo académico que incluya el acompañamiento del estudiante para que éste logre desarrollar habilidades académicas y de integración social, así como, el cultivo de un auto concepto positivo que lo guíe en el momento de enfrentar dificultades y que lo oriente en la manera de alcanzar sus metas.

2. Modelo de acompañamiento psicoeducativo y académico

Existen diversas perspectivas teóricas económicas, psicológicas, organizacionales y sociales para explicar la deserción de estudiantes universitarios (Braxton et al., 2000). Este estudio ha considerado la perspectiva organizacional de Tinto (1975, 1997), dada la experiencia de un programa inserto en una Institución de Educación Superior, y que por lo tanto debe articularse con los distintos servicios que ofrece a los estudiantes y que inciden en la deserción/ permanencia de éstos en la Educación Superior.

El modelo de Tinto (1975) es el que más ha influido en los estudios de deserción (Donoso y Schiefelbein, 2007; Saldaña y Barriga, 2010) y se sustenta mayoritariamente en los modelos de Bean (1980) y Spady (1970), otorgando importancia a factores externos a la universidad, así como a la integración que logra el estudiante con su ambiente académico y social. (Castaño et al., 2007, p. 55). Tinto propone que el estudiante ingresa a la universidad con una gama de atributos, experiencias y antecedentes familiares que afectan directa o indirecta en el desempeño del estudiante en la universidad.

Para ello, propone un modelo causal de cinco etapas: La primera etapa da cuenta de atributos previos al ingreso, donde se cuentan los antecedentes familiares, las características individuales y la escolaridad previa del alumno. La segunda etapa corresponde a las metas y compromisos del estudiante, relacionadas tanto con sus propias aspiraciones académicas como con la institución a la cual piensa ingresar.

La tercera etapa se refiere a las experiencias académicas y sociales vividas por el estudiante una vez dentro de la institución. La cuarta etapa se refiere la integración social y académica que logra el alumno a partir de las experiencias que ha vivido en la etapa anterior. Por último, en la quinta etapa existe un nuevo set de objetivos, metas y compromisos con la institución y con la educación, que

están mediados por el grado de integración que el estudiante alcanzó en el nivel anterior.

3. Modelos de duración a la problemática de la deserción

Distintos fenómenos que permiten ser descritos por alguna sucesión de cambios de estado, con diversidad de trayectorias y prolongaciones de la permanencia en los estados, pueden ser representados a través de modelos de duración. Los modelos de duración han sido aprovechados en diversas investigaciones aplicadas en ciencias sociales y de la salud.

Entre las más difundidas, se destacan modelos considerados para estudiar la duración del desempleo (Montero 2007; Arranz et al., 2000), la supervivencia ante la presencia de cierta enfermedad (Doménech 1992), el tiempo transcurrido hasta la primera concepción (Miranda 2005), o la demora de un graduado en conseguir su primer empleo (Salas 2007), entre otros. En estudios sobre deserción, la utilización de un modelo de duración implica poder considerar el efecto de diversas variables sobre la permanencia de alumnos hasta la ocurrencia de un abandono.

La deserción del alumno a lo largo del tiempo no es fácilmente identificable en el momento exacto en que se produce, sino que más bien pueden contarse los períodos en que el alumno desarrolló actividades académicas. Como se expuso previamente, la manera de especificar operativamente la deserción se circunscribe a identificar el cuatrimestre en el cual el alumno deja de inscribirse.

La aplicación de un modelo de duración a este problema permite estimar la probabilidad de que un alumno abandone sus estudios en cierto período, dado que no lo ha hecho hasta ese momento (a esta probabilidad condicional la denominamos función de riesgo en los modelos de duración). Los modelos de duración aplicados en trabajos similares (Giovanoli 2005; Lopera 2008; Castillo et

al., 2010; Castaño et al., 2007) suponen riesgo proporcional. Se asume así que las funciones de riesgo sólo difieren en un factor.

También es posible tener en cuenta los empates de los tiempos de duración, lo cual es una característica común en los datos debido a que el registro de la información se lleva a cabo en intervalos semestrales de tiempo (Castaño et al., 2007, 55). Por otra parte, los modelos de duración ofrecen herramientas para analizar diferentes tipos de decisiones que toman los estudiantes, no incluidos en los fenómenos que comúnmente se estudian: deserción y graduación.

2.2. Marco Institucional

1. Historia de la Facultad de Ingeniería de la Universidad de la Guajira

Según el ICFES (2002), son Universidades: Aquellas instituciones reconocidas actualmente como tales y las que acrediten su desempeño con criterio de universalidad en las siguientes actividades: en la investigación científica o tecnológica; la formación académica en profesiones o disciplinas y la producción, desarrollo y transmisión del conocimiento y de la cultura universal y nacional.

La Universidad de La Guajira, como producto de las Ordenanzas 011 y 012 de 1976 expedidas por la Asamblea Departamental y reglamentadas por el Decreto Gubernamental 523 de diciembre de 1976, fue creada como una entidad del orden departamental con personería jurídica, patrimonio propio y autonomía administrativa. Inició labores el 17 de febrero de 1977 con el nombre de UNIVERSIDAD EXPERIMENTAL DE LA GUAJIRA, ofreciendo el ciclo básico de Ingeniería Industrial (dándose así nacimiento a la FACULTAD DE INGENIERÍA) y de Administración de empresas. Posteriormente, por acuerdo N° 257 de Diciembre 19 de 1978, el Instituto Colombiano para el Fomento de la Educación Superior, le concedió licencia de iniciación de labores.

Según Valdivieso (2008), con la entrada en vigencia de la Ley 30 de 1992, la Universidad de la Guajira funciona como Institución Universitaria, hasta el 24 de mayo de 1995, cuando el Consejo Nacional de Educación Superior – CESU, mediante resolución 1770, emanada del MEN, le reconoce el estatus de Universidad, en atención a que había cumplido con la reglamentación vigente para ser reconocida como tal. Para cumplir con esta reglamentación, la Universidad de La Guajira presentó, a finales de 1994, el plan de desarrollo: “Un compromiso hacia el siglo XXI”, este plan orientó la gestión académica- administrativa, cualificando los recursos humanos y técnicos requeridos para su categorización.

En consecuencia, la historia de la FACULTAD DE INGENIERÍA está ligada al inicio mismo de la Universidad de La Guajira, de acuerdo con sus estatutos es una dependencia académica, adscrita a la Vicerrectoría Académica, en ella, están inscritos todos los programas académicos de formación profesional de pregrado cuyos ejes temáticos pertenecen, por definición, a los campos básicos de Ingeniería: Industrial, Ambiental, Sistemas, Civil y Mecánica y Salud ocupacional que fueron creados en los siguientes años:

1977 Programa Ingeniería Industrial nace con la fundación de la Universidad de La Guajira.

1993 Se creó el programa de Ingeniería del medio ambiente, posteriormente se cambió la denominación a Ingeniería Ambiental.

2001 Programa de Ingeniería de Sistemas y Telemática.

2008 Programa de Ingeniería Civil e Ingeniería Mecánica.

En Noviembre de 2012 el Consejo de Facultad acordó la apertura de cuatro programas profesionales cuyo origen es la Escuela Técnica, con estructuración programática por ciclos propedéuticos, es decir, los programas se ofrecerán con titulación en tres niveles: técnico, tecnológico y profesional. Dichos programas son: Ingenierías Agroindustrial, Minero, Electromecánica y seguridad e higiene

ocupacional. En marzo de 2013 se aprobó la ampliación de la oferta de Ingeniería de Sistemas a los Municipios de Maicao y Fonseca.

2. Programas de postgrados

Con relación a los programas de Posgrados, en el marco del Sistema de Universidades Estatales del Caribe SUE – Caribe la FACULTAD DE INGENIERÍA contó con Registro Calificado desde el año 2005 para la Maestría en Ciencias Ambientales, cuya primera cohorte inició labores en el segundo semestre de 2009 obtuvo renovación del registro mediante resolución No. 2581 del 14 de marzo de 2013.

Según el acuerdo 008 del 8 de Julio de 2010, fue creada la Maestría en Pedagogía de las Tecnologías de la Información y la Comunicación PTIC, de acuerdo a su estructura curricular aunque está inscrita en la Facultad de Ciencias de la Educación, recibe soporte técnico y humano de la Facultad de Ingeniería. De otro lado, están en proceso de creación la Maestrías en Producción y logística y la Maestría en Ingeniería Ambiental.

3. Estructura administrativa

Las dependencias académicas de La Universidad se denominan: Facultades, Programas, Escuelas, Institutos, Centros, Laboratorios, Grupos y otros. Según el artículo 50 del Estatuto General (2011), Facultad es la unidad básica de la organización académica que agrupa campos y disciplinas afines del conocimiento, profesores y personal administrativo, bienes y recursos; con el objeto de planificar, ofrecer y administrar programas curriculares de pregrado y postgrado, además, contribuir con la investigación y la extensión, de conformidad con las políticas y criterios emanados de los órganos de gobierno de la universidad.

La Facultad será dirigida por un Decano y un Consejo de Facultad. Cada Facultad será organizada administrativa y académicamente de acuerdo a sus propias características y de las funciones básicas de la Educación Superior: docencia, investigación y extensión. Tal como se observa en la figura, la Facultad, como órgano dependiente de la Vicerrectoría académica, se concibe como una estructura lineal en donde se propende por el trabajo en equipo, los niveles de autoridad son representados en línea vertical, formados; en donde el Decano representa la máxima autoridad, a continuación los demás niveles de la organización, pensados como equipos de trabajo que generen sinergia, logrando así la eficiencia y eficacia en la administración académico-administrativa de la Facultad.

4. Lineamientos estratégicos de la facultad

Francés (2006), los lineamientos estratégicos son los postulados fundamentales que plasman los principales aspectos estratégicos en una organización en particular, de acuerdo con las prácticas generalmente establecidas. Se clasifican en permanentes: Fines, misión, valores; semi permanentes: visión y políticas; y temporales: objetivos, indicadores, metas y estrategias.

5. Misión

La Facultad de Ingeniería forma profesionales en campos básicos de la ingeniería, permitiendo fortalecer la innovación y el desarrollo tecnológico, la generación de conocimiento en las diferentes áreas de dichos campos, de modo que, contribuyan a la solución de los grandes problemas del Departamento; lográndose mediante el aprovechamiento de los Recursos Naturales, enmarcados en la Sostenibilidad.

Se reconoce como parte de una región multicultural para lo cual diseña y desarrolla estrategias de Aprendizaje orientadas hacia la competitividad, eficiencia

y eficacia; con lo cual, busca proyectarse en un mundo globalizado. La investigación y la proyección social son parte fundamental de la formación integradora de sus egresados, sin desconocer la parte humana, ética y de sensibilidad por los problemas del entorno de los futuros profesionales de la Ingeniería.

Como estrategia para la formación en investigación de Talento Humano, la FACULTAD DE INGENIERÍA está ejecutando en Red (SUE-Caribe) el programa Maestría en Ciencias Ambientales y proponiendo otros programas de formación de posgrado y de alto nivel, Maestría y Doctorado, en diferentes áreas relacionadas con el quehacer de la Ingeniería.

6. Visión

Para el año 2018, la FACULTAD DE INGENIERÍA de la Universidad de La Guajira, será una unidad académica líder en la Costa Caribe Colombiana en la enseñanza de los campos básicos de las ingenierías; con el reconocimiento tanto local, nacional así como, internacional de sus programas a través de la acreditación de alta calidad. Ayudando a construir una sociedad más próspera, donde primen los derechos ciudadanos como la convivencia pacífica, la tolerancia y el respeto por el otro. La formación de sus egresados en pregrados y postgrados, estará marcada por el liderazgo, los conocimientos científicos, tecnológicos para la solución de los grandes problemas nacionales y la generación del desarrollo social y económico mediante la innovación.

7. Filosofía

La FACULTAD DE INGENIERÍA en la Universidad de La Guajira es un espacio que reconoce la diversidad cultural, fomenta la formación de los profesionales de la Ingeniería con espíritu: creativo, crítico, investigativo, innovador y social; valores que le permiten al egresado afrontar de manera responsable, ética, eficiente y

pertinente los retos de su entorno; sin dejar de lado su formación como persona y como profesional, buscando su integralidad en sus dimensiones como ser social, cultural y espiritual.

8. Principios

Reconociendo el valor rector de la universidad de La Guajira, La Honestidad (PEI, 2005), la FACULTAD DE INGENIERÍA de la universidad de La Guajira, asume como propios los siguientes principios y valores:

1. Desarrollo Humano

La FACULTAD DE INGENIERÍA, asume y propenderá por el Desarrollo Humano: “Es un paradigma de desarrollo que va mucho más allá del aumento o la disminución de los ingresos de un país. Comprende la creación de un entorno en el que las personas puedan desarrollar su máximo potencial y llevar adelante una vida productiva y creativa de acuerdo con sus necesidades e intereses. Las personas son la verdadera riqueza de las naciones. Por lo tanto, el desarrollo implica ampliar las oportunidades para que cada persona pueda vivir una vida que valore. El desarrollo es entonces mucho más que el crecimiento económico, que constituye sólo un medio si bien muy importante para que cada persona tenga más oportunidades. (PNUD, 2011).

2. Autonomía de aprendizaje, investigación, cátedra y expresión:

De acuerdo con el PEI (2005), la autonomía de aprendizaje, de investigación y de academia en la FACULTAD DE INGENIERÍA, queda establecida según el artículo 28 de la ley de 30 de 1992, que a la letra dice: “La autonomía universitaria consagrada en la Constitución Política de Colombia y de conformidad con la presente ley, reconoce a las universidades el derecho (...) a crear, organizar y

desarrollar sus programas académicos, definir y organizar sus labores formativas, académicas, docentes, científicas y culturales” (Ley 30 1992).

3. Excelencia en el desempeño académico

Formación universitaria acorde con las políticas de calidad académica propuestas por el Consejo Nacional de Acreditación – CNA, en el sistema universitario Colombiano, desarrollando al máximo su potencial intelectual, adquiriendo nuevos conocimientos y destrezas que lo capaciten para incorporarse al mundo intelectual y laboral en las mejores condiciones posibles o realizar tareas de investigación que redunden en beneficio de la Región o del País.

4. Transparencia

Considerada como la actitud y el comportamiento honestos de los actores de la comunidad universitaria, en concordancia con los principios y valores institucionales. (PEI, 2005).

5. Proactividad

La proactividad no significa sólo tomar la iniciativa, sino asumir la responsabilidad de hacer que las cosas sucedan; decidir en cada momento lo que queremos hacer y cómo lo vamos a hacer. En este sentido, la FACULTAD DE INGENIERÍA requiere de una comunidad académica dinámica, capaz de enrutarla hacia los mejores estándares de calidad académica.

6. Innovación

La FACULTAD DE INGENIERÍA, para su desarrollo a través del tiempo, requiere una dinámica orientada a fomentar su capacidad de innovación, ya que las organizaciones que incorporan la innovación a sus procesos y adoptan una actitud

abierta al cambio se posicionan mejor en el mundo competitivo, de modo que pueda adaptarse fácilmente a las dinámicas propias de los nuevos tiempos globalizadores, por ello, adopta la innovación como un principio regulador en su quehacer académico – administrativo, como un proceso de actualización permanente.

7. Espíritu investigativo

Puede entenderse, como espíritu investigativo el interés de la comunidad académica para aprender, razonar, pensar creativamente, tomar decisiones y buscar alternativas orientadas a resolver problemas, permitiendo el análisis y la comunicación de los resultados a la misma comunidad.

9. Valores

Fueron seleccionados los siguientes valores para la Facultad:

1. Compromiso: Se refiere al desarrollo de las capacidades y las posibilidades de cada miembro de la comunidad académica de manera que con su esfuerzo y compromiso pueda dar lo mejor de sí mismo, en aras de un proyecto de vida significativo y flexible. Este proceso está mediado por la interacción con los demás, y busca el bien tanto colectivo como el individual.

2. Eficiencia: Se refiere a la actitud frente al trabajo que lleva a las personas a realizar sus actividades con los mejores estándares de calidad, y la racionalización de los recursos utilizados.

3. Eficacia: Con este valor, la FACULTAD DE INGENIERÍA, propende por el cumplimiento de las actividades tanto académicas así como administrativas, se cumplan en igual o menor plazo que el estimado con antelación. Con ello, se busca la agilidad y prontitud en los procesos administrativos.

4. Competitividad: Según Francés (2006), hace referencia a las acciones que emprende una organización para cumplir con los objetivos propuestos en situaciones coyunturales del mercado, tales como la aparición de nuevos competidores o la incursión de un nuevo producto o servicio. La Facultad de Ingeniería busca ser creativa y crítica en su quehacer académico – administrativo para ofrecer un servicio educativo con altos estándares de calidad comparable en los contextos tanto Nacional como Internacional.

5. Responsabilidad: Ser responsable significa ser legal o éticamente capaz de rendir cuentas del cuidado o bienestar de otro; lo que implica la capacidad personal de rendir cuentas o la habilidad para actuar sin guía o autoridad superior. De acuerdo con lo anterior, una persona responsable toma decisiones conscientemente y acepta las consecuencias de sus actos, dispuesto a rendir cuenta de ellos, es decir, la virtud o disposición habitual de asumir las consecuencias de las propias decisiones, respondiendo de ellas ante alguien.

10. Políticas, acuerdos y reglamentos

La Facultad de Ingeniería se rige por las normas y reglamentos establecidos por la Universidad y, adicionalmente ha desarrollado algunos instructivos, políticas y reglamentos internos que se encuentran aprobados

2.3 Sistemas de Variables

DESERCIÓN ESTUDIANTIL: Aunque actualmente la definición de deserción estudiantil continúa en discusión, existe consenso en precizarla como un abandono que puede ser explicado por diferentes categorías de variables: socioeconómicas, individuales, institucionales y académicas. (MEN, 2009).

MINERÍA DE DATOS: La minería de datos puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. (Han et al., 2011, p. 8).

MINERÍA DE DATOS: Desde la perspectiva de la investigación, es una herramienta que funge significativamente en el volumen de datos categorizados por el conocimiento de los factores asociados a la deserción de estudiantes de instituciones de educación superior a fin de encontrar explicaciones y tomar decisiones acertadamente.

Cuadro 1. Matriz de Operacionalización de la Variable

Objetivo General: Evaluar el uso de la minería de datos para la estructura predictiva de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.				
OBJETIVOS ESPECÍFICOS	VARIABLE	DIMENSIONES	INDICADORES	
Caracterizar los factores determinantes de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.	DESERCIÓN ESTUDIANTIL	Factores determinantes	Individuales	
			Académicos	
			Institucionales	
			Socioeconómicos	
Identificar los recursos necesarios para la aplicación de la minería de datos en la Universidad de la Guajira.		MINERÍA DE DATOS	Recursos necesarios	Recursos Tecnológicos
				Recursos Humanos
Describir el modelo predictivo en minería de datos que se ajuste a la dinámica de la deserción estudiantil.	Modelo predictivo		Retención Acompañamiento Duración	
Proponer mediante la minería de datos un modelo para la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.	El logro de este objetivo se efectuará a través de los resultados de la presente investigación.			

Fuente: Propia del investigador, (2019).

3. MARCO METODOLOGICO

3.1 Enfoque

Desde el campo epistemológico porque aporta al entendimiento del desarrollo humano desde el paradigma positivista, a un nivel explicativo, lo cual supone un análisis descriptivo del fenómeno observado. Para comprender los factores determinantes de la deserción y la minería de datos se considera la perspectiva del desarrollo humano propuesto por Sen (2000), quien señala que la igualdad de las personas no se consigue sobre lo utilitario si no en la igualación de capacidades que permiten fortalecer las propias potencialidades, siendo el fin último el bienestar.

Es decir, todas las personas tienen la opción de desarrollar sus capacidades, pero existen factores externos e internos las que la afecta, bajo este parámetro se establece que la formación universitaria es una forma de desarrollo y por tanto existen factores determinantes de la deserción estudiantil y minería de datos que pueden detenerla.

Asimismo, se parte de la premisa que lo que conlleva a la igualdad es la posibilidad de formación universitaria, sin embargo, el acceso no es suficiente, sino que es necesario igualar capacidades y minar datos para realmente el estudiante tenga la opción de obtener formación profesional en condiciones de equidad e igualdad ya que solo así se conseguirá justicia social.

Por otra parte, las suposiciones sobre diseños y adelanto de modelos han comenzado a tolerado la técnica efectiva, esencialmente cuantitativa formulada por (Ramos, 2017), quien esboza que el enfoque de la investigación queda determinado por dos estándares paradigmáticas. Además, Hernández et al., (2010), señalan que el enfoque cuantitativo o paradigma positivista, ha germinado

como corriente de la ideología originada por las desiguales rutas en la indagación del conocimiento.

Adicionalmente, Álvarez et al. (2017), comentan que el enfoque o paradigma positivista ocupa una noción de la situación única, equitativa y visible, asimismo, puede ser dividida en partes para su estudio, lo que alcanza estar mostrando que los beneficios del investigador en este paradigma son los que establecen las dificultades de investigación y las conjeturas para su abordaje.

De la misma manera, Behar (2008), muestra que el enfoque cuantitativo recoge información empírica (de cosas o aspectos que se pueden contar, pesar o medir) y que por su naturaleza siempre arroja números como resultado, termina con datos numéricos y es fuerte en cuanto a la precisión del fenómeno mismo pero es débil en cuanto al papel del contexto o ambiente en la generación de esos datos.

Por otro lado, Hernández et al. (2014), hacen énfasis en que el enfoque cuantitativo usa la recolección de datos para probar hipótesis, con base en la medición numérica y el análisis estadístico, para establecer patrones de comportamiento y probar teorías; por lo antepuesto, son razones aptas para enmarcar la actual investigación con un enfoque positivista o cuantitativo al indagar evaluar el comportamiento de la variable Diseño de modelo en un contenido que pondera los hechos que han determinado la forma como se ha venido llevando el uso de la minería de datos para la estructura predictiva de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.

3.2 Tipo de investigación

De acuerdo a estudio, para el análisis o evaluación de la variable, se manejó la investigación de tipo descriptiva, que en relación con lo indicado por Supo (2015), quien señala que las investigaciones que detallan con una sola variable se

clasifican como descriptiva; de la misma manera la monografía se enmarca dentro del nivel de comprensión de investigación proyectiva de compromiso al propósito y los objetivos a alcanzar. La misma se enfoca en proponer mediante la minería de datos un modelo para la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira Sede Maicao.

Por otro lado, Hurtado (2012), añade que el tipo de investigación proyectiva pretende plantear soluciones a un ambiente determinado a partir de un proceso antepuesto de búsqueda que involucra indagar, describir y proponer opciones de cambios mas no obligatoriamente elaborar la proposición en el mismo orden. Sumado al contexto, se muestra que los tipos de estudios en el caso de las ciencias sociales, sostenidos en un modelo operativo, de un elemento de acción, están situados a suministrar respuestas o soluciones a problemas trazados en una determinada realidad: organizacional, social, económica, educativa, entre otras. (Balestrini 2006, p. 11).

Dado a los contextos anteriores se afina que, una investigación proyectiva es aquella que posee como propósito la creación de ideas, programas, diseños, modelos, entre otros; en el cual dicho logro nace de una insuficiencia y va propuesto a un área en específica. Dicha enunciación se concuerda a la doctrina de esta investigación, proponer mediante la minería de datos un modelo para la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira, dicho modelo tiene como finalidad optimizar la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira Sede Maicao quien ha verificado de acuerdo con Madrid, los docentes y la directiva del programa (2019), aquejada en la deserción y en consecuencia falta del minado de datos para poder predecir.

En relación a lo descrito con anterioridad, el entorno de la presente investigación se implanta dentro del juicio y características de un proyecto factible, pues tiene como objetivo facilitar solución a un problema existente. Por ello, Padilla (2013),

comenta que el proyecto factible consiste en una propuesta respaldada en un modelo operativo factible que ha de ser situada a resolver un problema diseñado o a satisfacer insuficiencias en una entidad o campo de interés territorial, bien nacional, o bien internacional.

Adicional a ello, Vásquez (2013), indica que el proyecto factible radica en la investigación, elaboración y desarrollo de una propuesta de un modelo operativo posible para solucionar problemas, imposiciones, o necesidades de organizaciones o grupos sociales. Puede relatar a la formulación de habilidades, tecnologías, métodos o procesos. El proyecto debe obtener apoyo en una investigación de tipo documental, de campo, o un diseño que incluya ambas características.

Por otra parte, en correspondencia a los conocimientos concernidos con el nivel de juicio o profundidad del estudio, se constituyó una investigación de nivel descriptiva, donde Méndez (2009), define que una vez identificada la ocurrencia de los hechos positivos resulta convertirlos en datos dispuestos de medidas y comparación; es decir, en los estudios de nivel descriptivo tanto el enfoque cuantitativo la estadística como el uso de la misma son retoques para concentrar y obtener información de la realidad observada. En esta investigación, los datos se obtendrán están relacionados con los factores determinantes, los recursos necesarios y los modelos predictivos para la minería de datos en el Programa de Ingeniería en Sistemas de la Universidad de la Guajira.

3.3 Diseño de la investigación

El término diseño se refiere al plan o estrategia concebida para obtener la información que se desea con el fin de responder al planteamiento del problema (Wentz, 2014; McLaren, 2014; Creswell, 2013a, Hernández et al., 2013 y Kalaian, 2008). Dado a ello, el diseño de investigación fue no experimental porque se realizan sin la manipulación deliberada de variables y en los que solo se observan

los fenómenos en su ambiente natural para después analizarlos. (Hernández et al., 2014, p. 152).

Además, la investigación no experimental según Mertens (2010), señala que es apropiada para variables que no pueden o deben ser manipuladas o resulta complicado hacerlo. Lo que se hace en la investigación no experimental es observar fenómenos tal como se dan en su contexto natural, para analizarlos. (The SAGE Glossary of the Social and Behavioral Sciences, 2009, p. 23). Del mismo modo, Sabino (2014), plantea que en los diseños de campo los datos primarios y de interés se recogen en forma directa de la realidad, mediante el trabajo concreto del investigador y su equipo.

Asimismo, el diseño es transeccional o transversal porque se recolectan datos en un solo momento, en un tiempo único. Su propósito es describir variables y analizar su incidencia e interrelación en un momento dado, también muestran que un diseño transversal o transaccional descriptivo tiene como objeto indagar la incidencia y los valores en que se manifiestan una o más variables dentro de un enfoque cuantitativo. (Hernández et al., 2014, p. 154). Seguidamente los diseños de investigación transeccional o transversal recolectan datos en un solo momento, en un tiempo único (Liu, 2008 y Tucker, 2004). Su propósito es describir variables y analizar su incidencia e interrelación en un momento dado. Es como “tomar una fotografía” de algo que sucede.

3.4 Población y muestra

3.4.1 Población

Según Tamayo y Tamayo (2007), evidencia que una población está determinada por sus características definitorias, por tanto es el conjunto de elementos que posea ésta característica se denomina población o universo. La población es la totalidad del fenómeno a estudiar en donde las unidades de población poseen una

característica común, la cual se estudia y da origen a los datos de la investigación. Igualmente, Arias (2012), expresa la población objetivo es un conjunto finito o infinito de elementos con características comunes para los cuales serán extensivas las conclusiones de la investigación.

Esta queda limitada por el problema y por los objetivos del estudio. La población finita es la agrupación en la que se conoce la cantidad de unidades que la integran. Además existe un registro documental de dichas unidades. Desde el punto de vista estadístico una población finita está integrada por número inferior a cien mil unidades. La población de acuerdo con Chávez (2007), es el universo de la investigación sobre la cual se pretende generalizar los resultados, y está constituida por características o estratos que permiten distinguir los sujetos unos de otros.

Por lo tanto, en el caso de la presente investigación las unidades de observación está constituido por 283 estudiantes del Programa de Ingeniería en Sistemas, por la Directiva y el personal de dicho programa, el cual se desglosa en 41 Profesores, 3 Ingenieros en Sistemas, 1 administrador de empresa, 1 técnico en informática, 1 Director principal del Programa y 1 Decana de facultad de ingeniería de la Universidad de la Guajira Sede Maicao, para un total de 331, la distribución es la siguiente:

Cuadro 2. Distribución de la población

UNIVERSIDAD DE LA GUAJIRA SEDE MAICAO		
PROGRAMA	CARGO	SUJETOS
INGENIERÍA EN SISTEMAS	Director principal	1
	Decana	1
	Ingenieros	3
	Administrador	1
	Técnico	1
	Profesores	41
	Estudiantes	283
TOTAL	331	

Fuente: Universidad de la Guajira Sede Maicao. (2019).

3.4.2. Muestra

En la presente investigación se aplicó el muestreo aleatorio simple, según Hernández et al. (2014), es un método de selección de n unidades sacadas de N, de tal manera que cada una de las muestras tienen la misma probabilidad de ser elegida. Adicionalmente, Arias, (2012), lo menciona como el procedimiento en el cual todos los elementos tienen la misma probabilidad de ser seleccionados. Para determinar el tamaño de la unidad de análisis se tomará en cuenta el tamaño de los conjuntos, por cuanto no excede de las 100.000 unidades, clasificándose como poblaciones finitas, hecho éste que incide en la selección de la fórmula estadística, para calcular la cantidad de individuos a encontrar, para lo cual en la presente investigación se utilizará la fórmula planteada por Sierra:

$$n = \frac{4 \times N \times P \times q}{E^2 (N \times 1) + 4 \times P \times q}$$

Dónde:

n = Tamaño de la Muestra

N = Tamaño de la Población

P y Q = Probabilidad de éxito o fracaso de los clientes con un valor de 95,5%, por lo que P y Q = 95,5.

E = Error máximo permisible por el investigador = 5%

4= Es constante por que se trabaja con un intervalo de confianza del 95%.

Sustituyendo la fórmula con los valores de los docentes, se obtiene una muestra de 22 usuarios:

$$N = \frac{(4) \times (41) \times (95,5) \times (95,5)}{(5)^2 (41 - 1) + (4) (95,5) \times (95,5)} = n = 22$$

Sustituyendo la fórmula con los valores de los estudiantes, se obtiene una muestra de 154 usuarios:

$$N = \frac{(4) \times (283) \times (95,5) \times (95,5)}{(5)^2 (283 - 1) + (4) (95,5) \times (95,5)} = n = 154$$

En base a lo antes señalado se muestra a continuación la muestra unidad de análisis:

Cuadro 3. Muestra

UNIVERSIDAD DE LA GUAJIRA SEDE MAICAO			
POBLACION	UNIDADES	PARTICIPACION	MUESTREO
Decana	1	0,30%	1
Director principal	1	0,30%	1
Ingenieros	3	0,91%	3
Administrador	1	0,30%	1
Técnicos	1	0,30%	1
Profesores	41	12,39%	22
Estudiantes	283	85,50%	154
TOTAL	331	100,00%	181

Fuente: Propia del investigador, (2019).

3.5 Técnica de recolección de datos

Para Hernández et al. (2014), la técnica de recolección de datos consiste en vías a través de los cuales el investigador registra datos observables que representa verdaderamente los conceptos o variables que se pretende medir. Asimismo Bavaresco (2009), indica que la técnica de recolección de datos constituye el conjunto de herramientas científicamente validadas por medio de los cuales se levanta los registros necesarios para comprobar un hecho o fenómeno en estudio.

En el presente estudio se utilizó la técnica de la encuesta definida por Hurtado (2010), como aquella técnica en la cual el investigador recurre a la aplicación de instrumentos de recolección de datos, para recabar la información requerida en función de los objetivos planteados y de observación, la cual es definida por

Hernández et al. (2014), como la técnica de investigación que consiste en observar personas, fenómenos, hechos, casos, objetos, acciones, situación, entre otros, con el fin de obtener determinada información necesaria para una investigación, la misma se realiza a través de un instrumento lista de cotejo la cual es un instrumento de evaluación o material de acciones que permiten obtener información relevante sobre el proceso de enseñanza y aprendizaje.

Por su parte, Méndez (2009), establece, que los cuestionarios son documentos estructurados o no que contiene un conjunto de reactivos, relativo a los indicadores de una variable y las alternativas de la respuestas, además expresa que a través de las encuesta se obtienen las opiniones de los individuos con relación a los objetos de la investigación.

A su vez, los instrumentos más utilizado para recolectar los datos es el cuestionario donde Hernández et al. (2014), consiste en un conjunto de pregunta respecto de una o más variable a medir. El proceso de recolección de datos en esta investigación fue realizado utilizando la técnica de la encuesta y como instrumento de recolección de datos, la encuesta tipo cuestionario el cual fue aplicado a las unidades de información haciendo uso de cuatro cuestionarios: tres cuestionarios auto administrado con escalamiento tipo Likert y de selección, para tres grupos de unidades: Directiva, Docentes y estudiantes; los mismos compuestos por: 21 ítems de clasificación de selección de alternativa para los estudiantes, otro de 12 ítems de tipo Likert con una ponderación de 5 alternativas de respuestas para la directiva y el ultimo de 11 ítems igualmente de tipo Likert para los docentes. (Ver anexos A, B y C).

Dentro de ese orden de ideas, las opciones de la escala likert como alternativa de repuesta a utilizar en cada ítems son cinco (5) opciones con dirección positiva, las cuales se encuentran representado de la siguiente manera: Muy de Acuerdo (MD) es positiva o favorable, a tal efecto la opción siempre recibe un puntaje de cinco (5), disminuyendo la puntuación en cada opción hasta llegar a la puntuación más

baja de un (1) punto para la menor que corresponde a Muy en Desacuerdo (MED).

A su vez, un último instrumento lista de cotejo el cual se encuentra estructurado por 4 ítems o preguntas relacionados a los indicadores recursos humanos y recursos tecnológicos con dos alternativas de respuestas de tipo cerrado Si y No para la selección de una. (Ver Anexo D).

Adicional a la descripción de los instrumentos de información, se aplicaron técnicas de extracción, transformación y carga de datos para convertir la información extraída de una base de datos transaccional hacia un formato apropiado para la aplicación de la técnica de minería de datos, obteniendo una data con las variables dependiente e independientes y sus dos categorías.

3.6 Técnicas para el procesamiento de la información

3.6.1 Presentación, análisis e interpretación de los datos (tratamiento estadístico)

Para la presentación y análisis de los datos, se utilizaron las herramientas Microsoft Excel, Árbol de decisión CHAID en SPSS. Se elaboraron tablas para analizar la relación entre las variables: factores individuales, académicos, institucionales y socioeconómicos. Los principales aspectos éticos relacionados con este tipo de investigación son las de privacidad respecto a la información personal tanto de los alumnos como docentes.

El uso de información que permita identificar, individualmente, a las personas requiere autorización y consentimiento previo al uso y publicación de la información. Para asegurar la privacidad de los datos se realizó un proceso para armonizar la información, asignando un identificador propio para cada registro en los datos y eliminar los atributos que permitieron identificar a las personas de los

diferentes registros como nombres, apellidos, dirección de domicilio, entre otros. De esta manera, se aseguró la validez de los resultados obtenidos en las pruebas sin exponer la información personal.

3.6.2. Recolección de data inicial

Se recolectó información de los instrumentos subidos a la web en plataforma de la Universidad de la Guajira Sede Maicao, para cada grupo de muestreo directiva, docentes y estudiantes, se extrajeron los siguientes datos: variables, seleccionando dos categorías correspondientes a los factores individuales, académicos, socioeconómicos e institucionales.

3.6.3 Exploración y verificación de la data inicial

La exploración inicial se efectuó para los datos de cada grupo de muestreo y para cada factor determinante de la deserción: Individuales, académicos, socioeconómicos e institucionales, luego los recursos y los modelos predictivos. Se halló, en la información de la que han sido extraídos, de la base de datos, hacia las hojas de cálculo de Microsoft Excel ignoran los espacios en blanco al momento de mostrar la información, lo que no ocurre en las bases de datos y en programas de análisis como SPSS, por lo que fue necesario, en la siguiente tarea, realizar la eliminación de los espacios y reemplazar los caracteres que se consideren erróneos.

3.6.4 Preparación y selección de la data

Una vez realizada la verificación de la data, en la fase anterior, se integró la información en un solo conjunto de datos. Se descarta información incompleta, faltante o incongruente para evitar que se analicen datos que puedan crear errores al momento de ejecutar la técnica o modelo de minería de datos. Se generó una variable para cada columna, en la tabla, a fin de diferenciar los datos y también

eliminar la dependencia de otros identificadores que pudieran existir en la data, en este caso, el código del alumno.

3.6.5 Limpieza de datos

En esta etapa, se aplicaron las funciones de EXCEL para eliminar los espacios en blanco existentes en los datos de tipo texto y se aplicaron las funciones de redondeo para los valores numéricos con decimales. Se eliminaron las informaciones personales que pudieran identificar, individualmente, a cada alumno como código del alumno, apellido paterno, apellido materno, nombre y dirección.

3.6.6 Construcción de la data

Esta fase se inició con la construcción de nuevas características para enriquecer la data con la finalidad de adaptarla a que sea apta para la aplicación de minería de datos. Entre las características agregadas se hallan:

1. RANGO DE EDAD
2. GENERO
3. ESTADO CIVIL
4. POSICION ENTRE LOS HERMANOS
5. APEGO FAMILIAR
6. VALORES CONSOLIDADOS
7. SALUD
8. APOYO ECONOMICAMENTE DE LOS PADRES EN LA FORMACION ACADEMICA
9. APOYO EDUCATIVAMENTE DE LOS PADRES EN LA FORMACION ACADEMICA
10. ESTRATO ECONOMICO
11. SITUACION LABORAL
12. NIVEL DE ESTUDIOS DE PADRES

13. DESERCIÓN

3.6.7 Minería de datos

En una primera iteración para detectar posibles errores en la data, se realizó el estudio con los ingresantes del segundo semestre, 2018-2. Se diseñó el Árbol de decisiones CHAID en SPSS: Para crear el modelo basado en la técnica de árbol de decisiones que permitió establecer un conjunto de reglas para poder realizar la predicción y analizar la importancia de cada variable independiente utilizada.

4. RESULTADOS DE LA INVESTIGACIÓN

El presente capítulo corresponde al análisis y discusión de los resultados obtenidos del proceso de recolección de información, mediante la estadística descriptiva - proyectiva, estableciéndose las frecuencias, porcentajes, medias y desviación estándar de éstos, y mostrar siguiendo el precepto de exposición de las variables: Deserción estudiantil y Minería de Datos.

4.1 Análisis de los resultados

Para analizar los datos recolectados a través de la aplicación de los instrumentos elaborado por el autor de la presente investigación, (Ver anexos A,B,C y D), se utilizó la estadística descriptiva - proyectiva, reseñando la información a través de cuadros indicando las frecuencias absolutas acompañadas de los respectivos porcentajes con el promedio del total de los datos, como la media y desviación para señalar la calificación que permitirá realizar la respectiva interpretación como el diseño y el análisis del modelo de árbol de decisión CHAID en SPSS.

Los cuales indican a los objetivos específicos esbozados en relación con el objetivo general, a fin de Evaluar el uso de la minería de datos para la estructura predictiva de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira, es para exponer elementos que suministren la reconstrucción teórica del análisis de las variables de la investigación.

OBJETIVOS ESPECIFICOS

Caracterizar los factores determinantes de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira.

Variable: Deserción estudiantil

Dimensión: Factores determinantes

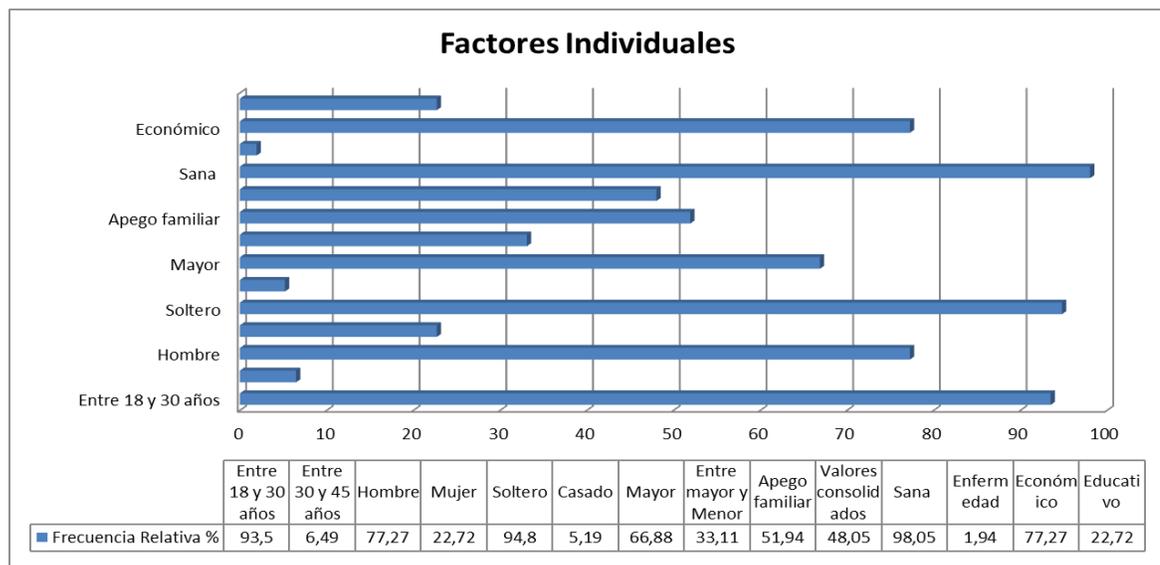
Indicador: Factores Individuales

Grupo de Muestreo: Estudiantes

Tabla 1. Factores Individuales

INDICADOR	FACTORES INDIVIDUALES													
	Ítem 1		Ítem 2		Ítem 3		Ítem 4		Ítem 5		Ítem 6		Ítem 7	
ALTERNATIVAS DE RESPUESTAS	Entre 18 y 30 años	Entre 30 y 45 años	Hombre	Mujer	Soltero	Casado	Mayor	Entre mayor y Menor	Apego familiar	Valores consolidados	Sana	Enfermedad	Económico	Educativo
Frecuencia Relativa %	93,50	6,49	77,27	22,72	94,80	5,19	66,88	33,11	51,94	48,05	98,05	1,94	77,27	22,72
TOTAL	100		100		100		100		100		100		100	
MEDIA DEL INDICADOR	49,99													
CATEGORÍA DEL INDICADOR	ALTO NIVEL													
DESVIACION ESTANDAR DEL INDICADOR	35,05													
CATEGORÍA DEL INDICADOR	ALTO NIVEL													

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Grafico 1. Frecuencia Relativa sobre los Factores Individuales

Según los datos mostrados en el gráfico N° 1, se evidencia que en ítem 1 del rango de edad el 93,5% de los estudiantes posee edades entre 18 y 30 años, además, en el ítem 2 genero, un 77,27% son hombres y un 22,72% son mujeres, en el ítem 3 estado civil, se observó que el 94% de los estudiantes son solteros, en el ítem 4 posición entre hermanos un 66,88% son el mayor de sus hermanos, en el ítem 5 prevalencia al hogar se mostró que el 51,94% posee apego familiar y un 48,05% posee valores consolidados.

Asimismo, en el ítem 6 salud se evidencia que el 98,5% se encuentra sano, en el ítem 7 en relación al apoyo de los padres en su formación escolar se reveló que un 77,27% sus padres los apoyan económicamente y 22,72% los apoyo educativamente, todos estos resultados no afectan la deserción estudiantil debido a que la desviación estándar de este indicador factores individuales fue de 35,05 clasificándola como una alta desviación lo cual se interpreta como nula dependencia de los factores individuales con la variable deserción estudiantil.

Estos resultados presentan controversias a lo planteado por Spady (1970), Tinto (1975), Ethington, (1970), Giovagnoli, (2002) menciona al respecto factores como la personalidad, dificultades de adaptación personal a la Institución, incompatibilidad entre la vida académica y las exigencias del trabajo, desencanto entre la vida académica y las exigencias del trabajo, dificultad en la relación enseñanza-aprendizaje, añadiendo otros factores como la edad, género y estado civil, posición entre los hermanos, entorno familiar, calamidad y problemas de salud, integración social.

De igual manera, Spady (1970), Tinto (1975), Ethington, (1970), Giovagnoli, (2002), los cuales han confirmado que los mismos determinan un alto porcentaje la deserción estudiantil, lo cual muestra controversias a lo observado en el Sistema de Información Académico (SIA) los resultados recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao.

Variable: Deserción estudiantil

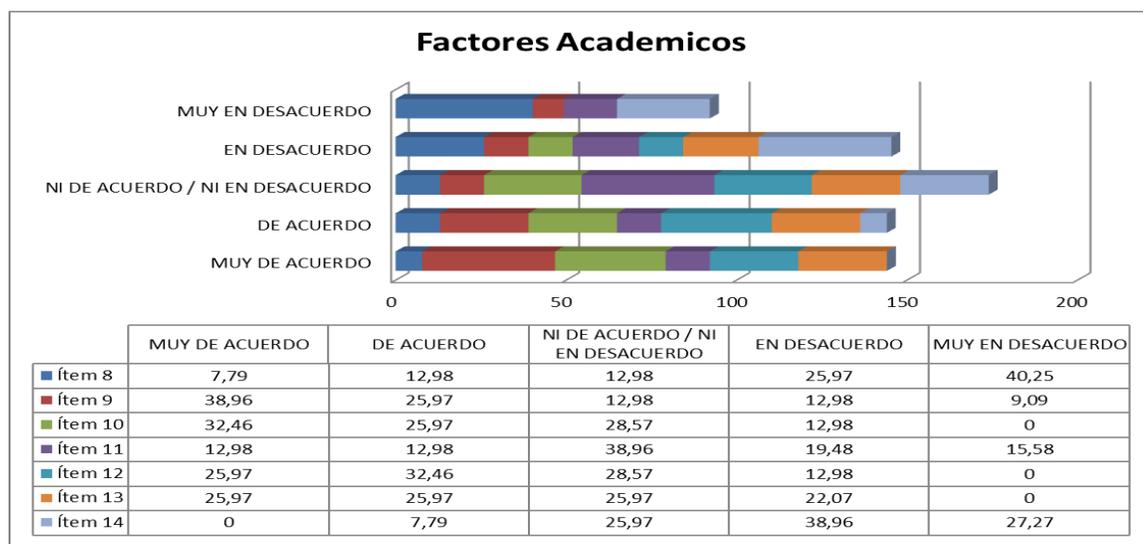
Dimensión: Factores determinantes; Indicador: Factores Académicos

Grupo de Muestreo: Estudiantes

Tabla 2. Factores Académicos

INDICADOR	FACTORES ACADEMICOS													
	Ítem 8		Ítem 9		Ítem 10		Ítem 11		Ítem 12		Ítem 13		Ítem 14	
ALTERNATIVAS DE RESPUESTAS	Fa	Fr %	Fa	%	Fa	Fr %	Fa	Fr %	Fa	Fr %	Fa	Fr %	Fa	Fr %
MUY DE ACUERDO	12	7,79	60	38,96	50	32,46	20	12,98	40	25,97	40	25,97	0	0
DE ACUERDO	20	12,98	40	25,97	40	25,97	20	12,98	50	32,46	40	25,97	12	7,79
NI DE ACUERDO / NI EN DESACUERDO	20	12,98	20	12,98	44	28,57	60	38,96	44	28,57	40	25,97	40	25,97
EN DESACUERDO	40	25,97	20	12,98	20	12,98	30	19,48	20	12,98	34	22,07	60	38,96
MUY EN DESACUERDO	62	40,25	14	9,09	0	0	24	15,58	0	0	0	0	42	27,27
TOTAL	154	100	154	100	154	100	154	100	154	100	154	100	154	100
MEDIA DEL INDICADOR	22,07													
CATEGORÍA DEL INDICADOR	MODERADO NIVEL													
DESVIACION ESTANDAR DEL INDICADOR	11,78													
CATEGORÍA DEL INDICADOR	MODERADO NIVEL													

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Grafico 2. Frecuencia Relativa sobre los factores académicos

En el gráfico n° 2, se presenta la discusión de los ítems pertenecientes a el indicador factores académicos en el grupo de muestreo de los estudiantes, la cual responde a la variable: deserción estudiantil, donde se aprecia el número de sujetos válidos para el estudio el cual es 154, aportando medias específicas para cada alternativa de respuesta las cuales a su vez reflejaron el desarrollo de la investigación. Por tanto estos resultados se obtuvieron gracias a las respuestas de cada uno de los sujetos.

Para el ítems N° 8, cuyas alternativas de respuestas fueron: que el 40,25% de la población manifestaron que muy en desacuerdo ingresó al Programa de Ingeniería en Sistemas por efecto de obtener un cupo universitario para luego cambiarse de carrera, lo cual ratifica un nivel moderado, que representa variaciones como existencia de un gran número de estudiantes que utilizan la carrera de trampolín para cambiarse a otra carrera.

De igual modo, se procedió a establecer con ayuda de la estadística descriptiva para el ítem N° 9, como se menciona a continuación: un 38,96% que muy de acuerdo que el estudiante dedica tiempo suficiente a la programación e investigación del Programa de Ingeniería en Sistemas, lo cual indica desajustes en el programa, debido a que existe un numero considerado que no le dedica el tiempo necesario al programa.

Ahora bien para el ítem N° 10, se evidenció que un 32,46% de la población señaló que muy de acuerdo, existe buena relación con los docentes del programa, la cual se ubica en la categoría de moderado nivel, provocando desequilibrio en el programa debido a que existe numerosa población estudiantil que muy poco posee buena relación con los docentes.

Por otro lado, para el ítems N° 11, se comprobó que un 38,96% de la población señaló que ni de acuerdo, ni en desacuerdo, que las exigencias del Programa de Ingeniería en Sistemas causan deserción, la cual se ubica en la categoría de

moderado nivel, provocando un desajuste dado a que la mayoría de los estudiantes no se encuentra seguro que las exigencias del programa causan o no deserción en el mismo, causando un inquietud por parte de los docentes y directiva del programa.

De igual forma, se procedió a establecer con ayuda de la estadística descriptiva para el ítem N° 12, como se menciona a continuación: un 32,46% que de acuerdo que el contenido programático es suficiente para el Programa de Ingeniería en Sistemas, lo cual significa que menos de la mitad de los estudiantes lo consideran y otros no, lo que genera desequilibrio en la implementación o modificación del contenido programático.

Por consiguiente, para el ítem N° 13, se observó que un 25,97% de la población señaló que muy de acuerdo, de acuerdo y ni de acuerdo / ni en desacuerdo, posee un alto nivel de satisfacción sobre el Programa de ingeniería en sistemas, la cual se ubica el desajuste en las respuestas ocasionando un desequilibrio en el nivel de satisfacción que tienen algunos estudiantes por el programa, y por ende busca a la directiva indagar las causas del porque el alumnado no posee satisfacción del programa.

Para el ítem N° 14, cuyas alternativas de respuestas fueron: que el 38,96% de la población manifestaron que muy en desacuerdo que piensa desertar al Programa de Ingeniería en Sistemas, lo cual certifica un nivel moderado, que representa desviaciones como presencia de un gran número de estudiantes que piensa desertar del programa, generando falencias al programa por la poca población inscrita.

Todo estos resultados, difieren que los factores académicos en relación a los estudiantes afectan la deserción de los mismos al programa de ingeniería en sistemas de la Universidad de la Guajira Sede Maicao, estos resultados muestran controversias a lo planteado Castaño et al. (2008), en el ejemplo del grado de

satisfacción del estudiante con el programa parece ser una variable que incide directamente en el riesgo de desertar, al encontrarse que aquellos estudiantes con un nivel de satisfacción alto o medio tienen menor riesgo de desertar en comparación con quienes tienen un bajo grado de satisfacción.

Variable: Deserción estudiantil

Dimensión: Factores determinantes

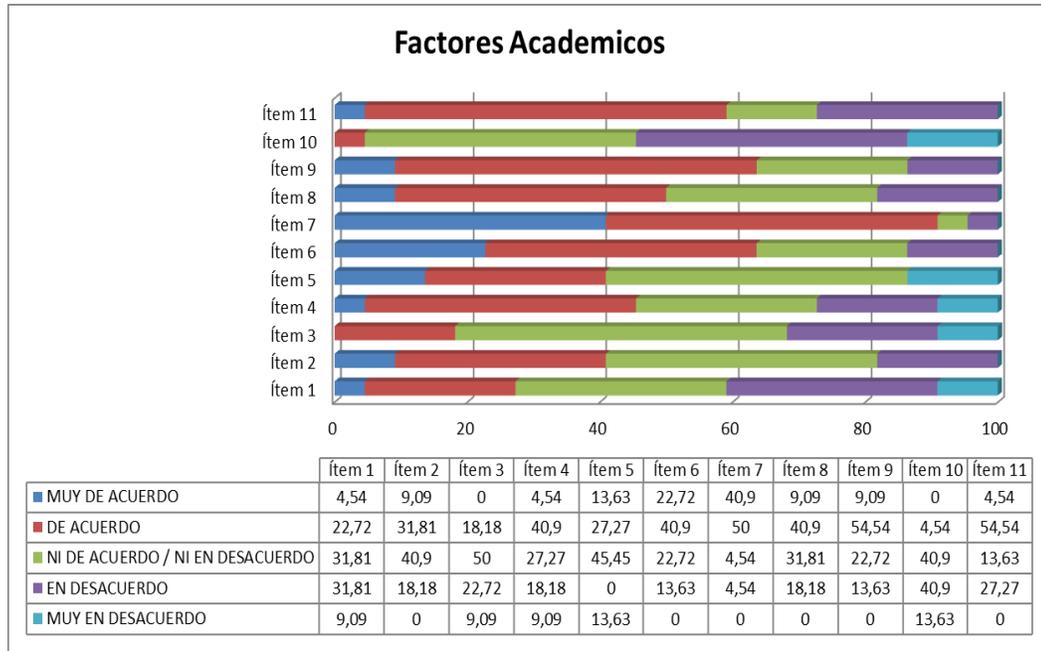
Indicador: Factores Académicos

Grupo de Muestreo: Docentes

Tabla 3. Factores Académicos

INDICADOR	FACTORES ACADEMICOS																					
	Ítems 1		Ítems 2		Ítems 3		Ítems 4		Ítems 5		Ítems 6		Ítems 7		Ítems 8		Ítems 9		Ítems 10		Ítems 11	
	Fa	Fr %	Fa	Fr%	Fa	Fr %	Fa	Fr%	Fa	Fr %	Fa	Fr %	Fa	Fr %								
MUY DE ACUERDO	1	4,54	2	9,09	0	0	1	4,54	3	13,63	5	22,72	9	40,90	2	9,09	2	9,09	0	0	1	4,54
DE ACUERDO	5	22,72	7	31,81	4	18,18	9	40,90	6	27,27	9	40,90	11	50	9	40,90	12	54,54	1	4,54	12	54,54
NI DE ACUERDO / NI EN DESACUERDO	7	31,81	9	40,90	11	50	6	27,27	10	45,45	5	22,72	1	4,54	7	31,81	5	22,72	9	40,90	3	13,63
EN DESACUERDO	7	31,81	4	18,18	5	22,72	4	18,18	0	0	3	13,63	1	4,54	4	18,18	3	13,63	9	40,90	6	27,27
MUY EN DESACUERDO	2	9,09	0	0	2	9,09	2	9,09	3	13,63	0	0	0	0	0	0	0	0	3	13,63	0	0
TOTAL	22	100	22	100	22	100	22	100	22	100	22	100	22	100	22	100	22	100	22	100	22	100
MEDIA DEL INDICADOR	18,18																					
CATEGORÍA DEL INDICADOR	MODERADO NIVEL																					
DESVIACION ESTANDAR DEL INDICADOR	16,47																					
CATEGORÍA DEL INDICADOR	MODERADO NIVEL																					

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Grafico 3. Frecuencia Relativa sobre los factores académicos

En el gráfico n° 3, se presenta la discusión de los ítems del indicador factores académicos del grupo de muestreo de docentes, la cual responde a la variable: deserción estudiantil, donde se aprecia el número de sujetos válidos para el estudio el cual es de 22, aportando media y desviación estándar específica para el indicador.

Por tanto estos resultados se obtuvieron gracias a las respuestas de cada uno de los sujetos. Para el ítem 1, se evidenció que un 31,81 % de la población manifestaron que ni de acuerdo, ni en desacuerdo, como en desacuerdo que los estudiantes poseen conocimientos previos sobre el programa de ingeniería en sistemas, un 22,72% de acuerdo, un 9,09% de muy desacuerdo y un 4,54% muy de acuerdo, lo cual indica que los docentes muy poco poseen conocimiento si los estudiantes conocen todo lo relacionado al programa de ingeniería en sistemas.

Por otro lado, para el ítem 2, se mostró que un 40,9% respondió que ni acuerdo, ni en desacuerdo conocen si el ingreso de los estudiantes al programa es por vocación, además un 31,81% de acuerdo, un 18,18% en desacuerdo, un 9,09% muy de acuerdo y un 0% muy en desacuerdo, lo cual muestra que los docentes muy pose se interesan si el estudiante ingresa al programa por vocación.

Consecuentemente, para el ítem 3 se observó que el 50% de los docentes respondieron que ni de acuerdo, ni en desacuerdo, conocen si el ingreso del estudiante al programa es por falta de oportunidades, asimismo, un 22,72% en desacuerdo, un 18,08% de acuerdo, un 9,09% muy en desacuerdo y un 0% muy de acuerdo, lo cual indica que los docentes no muestran interés en conocer las razones por las cuales el estudiante ingresa al programa de ingeniería en sistema.

De igual manera, para el ítem 4, se evidencia que un 40,9% de la población manifestaron que de acuerdo conocen que el ingreso del estudiante al programa es por falta de recursos económicos, a su vez, un 27,27% ni de acuerdo, ni en desacuerdo, un 18,18% en desacuerdo, un 9,09% muy en desacuerdo y un 4,54% muy de acuerdo, lo cual muestra que los docentes conocen en algunas oportunidades que los estudiantes ingresan o no al programa por la falta económica.

En relación al ítem 5, se observó que el 45,45% de los docentes respondieron que ni de acuerdo, ni en desacuerdo conocen si los estudiantes del Programa presentan inconvenientes para asistir a clases en el horario establecido, asimismo, un 27,27 % de acuerdo, un 13,63% que muy de acuerdo y muy en desacuerdo y un 0% en desacuerdo, lo cual indica que los docentes que poco conocen si los estudiantes presentan inconvenientes para asistir al horario de clases del programa.

De igual forma, para el ítem 6, se evidencia que el 40,9% manifestó que de acuerdo que la universidad beneficia a los estudiantes con el servicio de rutas

estudiantiles urbanas e intra – urbanas, a su vez, un 22,72% que muy de acuerdo y ni de acuerdo, ni en desacuerdo, un 13,63% en desacuerdo y un 0% muy de acuerdo, lo cual muestra que los docentes conocen que la universidad le ofrece las rutas estudiantiles a su población.

Consecuentemente, para el ítem 7, se observó que un 50% respondió que de acuerdo identifica por medio de observaciones físicas y mentales cuales son estudiantes desertores, asimismo, un 40,9% muy de acuerdo, un 4,54% ni de acuerdo, ni en desacuerdo y un 0% muy en desacuerdo, lo cual indica que varias oportunidades los docentes evidencia gracias a lo físico y las actitudes que el estudiante piensa en desertar.

Sin embargo, para el ítem 8, se observó que el 40,9% de los docentes respondió que de acuerdo en relación a que los aspectos económicos influyen en la deserción estudiantil del programa de ingeniería en sistemas, a su vez, un 31,81% en ni de acuerdo, ni en desacuerdo, un 18,98% en desacuerdo, un 9,09% de acuerdo y un 0% en desacuerdo, lo cual muestra que los docentes en ciertas oportunidades ha evidenciado que los aspectos económico influye a que el estudiante deserte del programa de ingeniería en sistemas.

Por otro lado, para el ítem 9, se observó que el 54,54% manifestó que de acuerdo los estudiantes manifiestan poco hábitos de estudios en los momentos antes y después del acto de clase, asimismo, un 22,72% manifiesta que ni de acuerdo, ni en desacuerdo, un 13,63% en desacuerdo, 9,09% muy de acuerdo y un 0% muy en desacuerdo, lo cual muestra que los docentes conocen los pocos hábitos de estudios que demuestran los estudiantes antes y después de clases.

De igual forma, para el ítem 10, se muestra que el 40,9% respondieran que ni de acuerdo, ni en desacuerdo y en desacuerdo existe compromiso educativo al programa por parte de los estudiantes, asimismo, un 13,63% muy en desacuerdo, un 4,54% de acuerdo y un 0% muy de acuerdo, lo cual indica que muy poco

conocen si los estudiantes poseen compromiso al programa de ingeniería en sistemas.

Por último, para el ítem 11, se evidenció que el 54,54% de los docentes respondieron que de acuerdo, que los estudiantes del programa de ingeniería en sistemas presentan motivación en la clase, además, un 27,27% en desacuerdo, un 13,63% ni de acuerdo, ni en desacuerdo, un 4,54% muy de acuerdo y un 0% muy en desacuerdo, lo cual muestra que los estudiantes en la mayoría de los casos muestran motivación de sus clases.

Debido a todos los resultados mostrados, se observan similitudes a lo planteado por Castaño et al., (2007), quienes muestran que el grado de satisfacción del estudiante con el programa parece ser una variable que incide directamente en el riesgo de desertar, al encontrarse que aquellos estudiantes con un nivel de satisfacción alto o medio tienen menor riesgo de desertar en comparación con quienes tienen un bajo grado de satisfacción.

De igual manera dentro del conjunto de variables académicas, las que parecen ser importantes en la explicación del riesgo de deserción son: el desempeño académico en la universidad. Los resultados muestran que a mayor número de créditos cursados, menor es el riesgo de desertar, mientras a mayor número de créditos reprobados, es decir a mayor repetida, mayor es el riesgo.

Variable: Deserción estudiantil

Dimensión: Factores determinantes

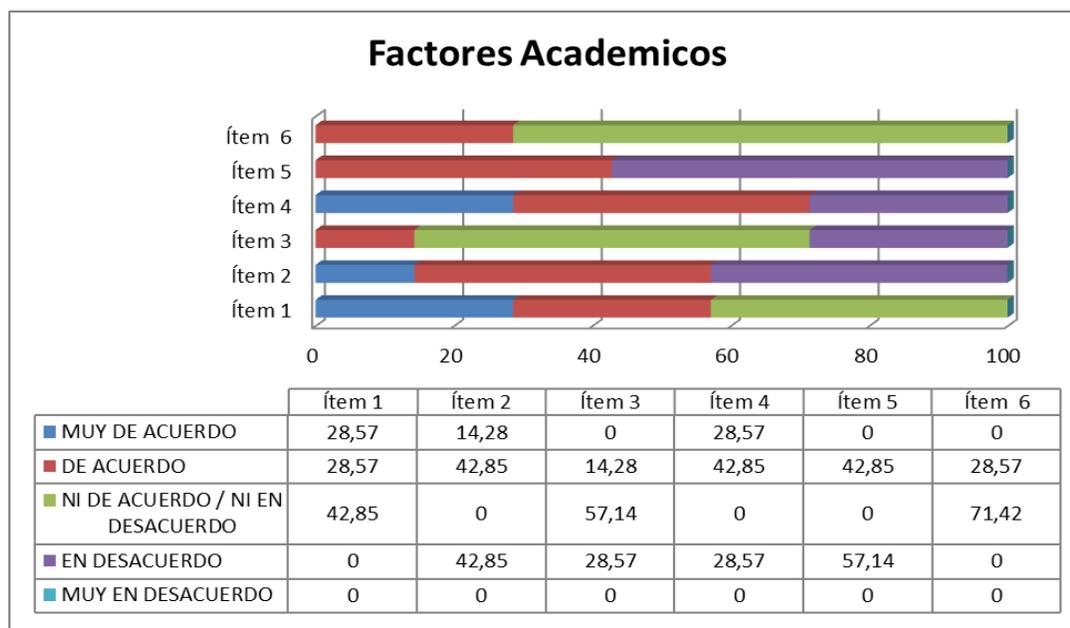
Indicador: Factores Académicos

Grupo de Muestreo: Directiva

Tabla 4. Factores Académicos

INDICADOR	FACTORES ACADÉMICOS											
	Ítems 1		Ítems 2		Ítems 3		Ítems 4		Ítems 5		Ítems 6	
	Fa	Fr %	Fa	%	Fa	Fr %	Fa	Fr %	Fa	Fr %	Fa	Fr %
MUY DE ACUERDO	2	28,57	1	14,28	0	0	2	28,57	0	0	0	0
DE ACUERDO	2	28,57	3	42,85	1	14,28	3	42,85	3	42,85	2	28,57
NI DE ACUERDO / NI EN DESACUERDO	3	42,85	0	0	4	57,14	0	0	0	0	5	71,42
EN DESACUERDO	0	0	3	42,85	2	28,57	2	28,57	4	57,14	0	0
MUY EN DESACUERDO	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	7	100	7	100	7	100	7	100	7	100	7	100
MEDIA DEL INDICADOR	14,28											
CATEGORÍA DEL INDICADOR	MODERADO NIVEL											
DESVIACION ESTÁNDAR DEL INDICADOR	22,06											
CATEGORÍA DEL INDICADOR	MODERADO NIVEL											

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Grafico 4. Frecuencia Relativa sobre los factores académicos

En relación a lo observado en el gráfico N° 4, se muestra para el ítem 1 el 42,85% de la directiva manifestó que el Programa de Ingeniería en Sistemas posee un alto índice de estudiantes reprobados, asimismo, un 28,57% que de acuerdo y que

muy de acuerdo, un 0% que en desacuerdo y muy en desacuerdo, lo cual indica que la directiva posee conocimiento que el programa posee un alto porcentaje de deserción estudiantil.

Asimismo, para el ítem 2, se observó que el 42,85% de la directiva manifestó que de acuerdo y en desacuerdo, conocen que el alumnado posee conocimiento sobre el perfil de ingreso del programa de ingeniería en sistemas, asimismo, un 14,28% que manifestó que muy de acuerdo y un 0% que ni de acuerdo, ni en desacuerdo como muy en desacuerdo, lo cual muestra que la directiva en ciertas ocasiones observa que el alumnado posee conocimientos sobre el perfil del programa.

Por otro lado para el ítem 3, se evidenció que 57,14% de la directiva manifestó que ni de acuerdo, ni en desacuerdo conocen que los estudiantes dedican horas suficientes al estudio aun cuando conocen los niveles de exigencia de la carrera, además un 28,57% en desacuerdo, un 14,28% de acuerdo y un 0% de muy de acuerdo y muy en desacuerdo, lo cual muestra que la directiva en la mayoría de los casos poco conoce si los estudiantes dedican las horas suficientes al estudio del contenido programático del programa de ingeniería en sistemas.

Para el ítem 4, se muestra que el 42,85% de la directiva respondieron que de acuerdo conocen que el Programa de Ingeniería en Sistemas posee una alta calidad, de igual manera, un 28,57% muy de acuerdo y en desacuerdo y un 0% de ni de acuerdo, ni en desacuerdo como muy en desacuerdo, lo cual indica que la directiva en la mayoría estar al tanto que el programa de ingeniería en sistemas posee una alta calidad.

Seguidamente, el ítem 5, mostró que el 57,14% de la directiva respondió que está en desacuerdo que el programa de Ingeniería en Sistemas ejecuta una prueba de admisión para el ingreso al mismo, asimismo, un 42,85% de acuerdo y un 0% que muy de acuerdo, ni de acuerdo, ni en desacuerdo, como muy en desacuerdo, lo

cual muestra que la directiva en ciertas oportunidades enuncia que el programa en ingeniería en sistemas ejecuta una prueba de admisión para el ingreso al mismo.

Por último, para el ítem 6, se evidencia que el 71,42% de la directiva mostró que ni de acuerdo, ni en desacuerdo el Programa de Ingeniería en Sistemas posee una base de datos de estudiantes que ingresan, desertan y egresan, asimismo, un 28,57% de acuerdo y un 0% muy de acuerdo, en desacuerdo y muy en desacuerdo, lo cual conduce a que la directiva conoce poco sobre la base de datos que posee el programa de ingeniería en sistemas.

Dado a estos resultados presentan controversias a lo planteado por Báez y Pedraza, (2011), quienes señalan que desde la orilla de lo académico emerge otro factor decisivo en la deserción estudiantil. Se alude tanto a los vacíos en conocimiento, habilidades cognitivas, problemas de aprendizaje, deficiencias en los niveles de lectura inferencial y crítica, como de escritura y desarrollo del pensamiento lógico-matemático, como a la ausencia de hábitos de estudio, tipo de colegio, calidad del programa, método de estudio, número de materia, insatisfacción con el programa, disciplina, entusiasmo, compromiso, exigencia, perseverancia, motivación para aprender, entre otros. Y, yendo aún más allá, a la carencia de estilos de vida que propulsen el éxito académico.

De igual forma, Báez y Pedraza, (2011), señalan que los factores académicos han permitido que los mismos determinan un alto porcentaje en la determinación de la deserción estudiantil, lo cual pauta polémicas a lo observado en el Sistema de Información Académico (SIA) resultados recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao.

Variable: Deserción estudiantil

Dimensión: Factores determinantes

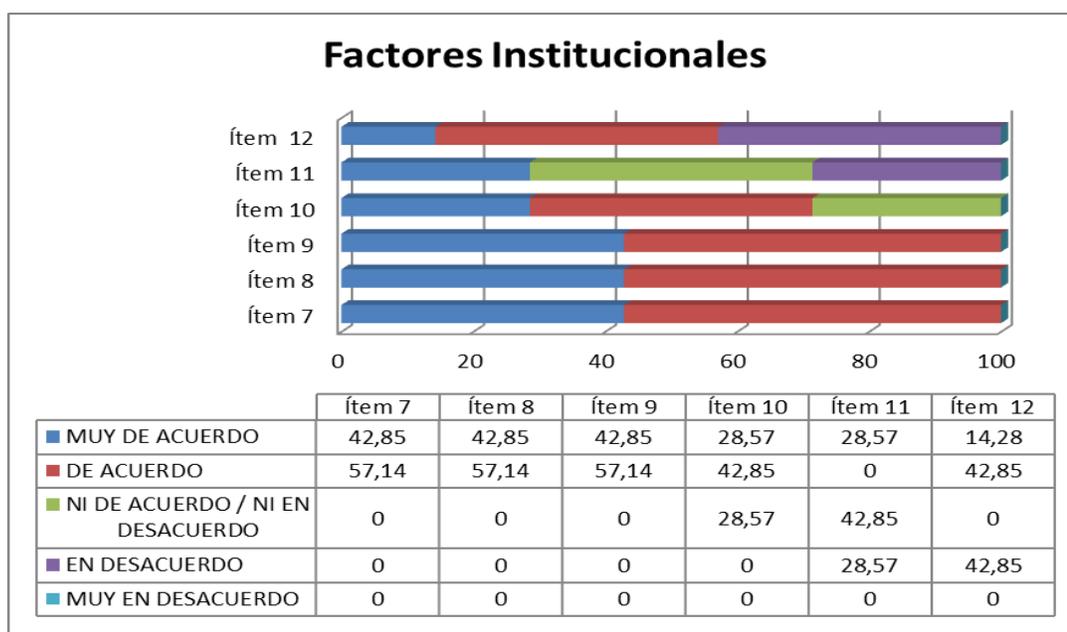
Indicador: Factores institucionales

Grupo de Muestra: Directiva

Tabla 5. Factores Institucionales

INDICADOR	FACTORES INSTITUCIONALES											
	Ítems 1		Ítems 2		Ítems 3		Ítems 4		Ítems 5		Ítems 6	
	Fa	Fr %	Fa	%	Fa	Fr %	Fa	Fr %	Fa	Fr %	Fa	Fr %
MUY DE ACUERDO	3	42,85	3	42,85	3	42,85	2	28,57	2	28,57	1	14,28
DE ACUERDO	4	57,14	4	57,14	4	57,14	3	42,85	0	0	3	42,85
NI DE ACUERDO / NI EN DESACUERDO	0	0	0	0	0	0	2	28,57	3	42,85	0	0
EN DESACUERDO	0	0	0	0	0	0	0	0	2	28,57	3	42,85
MUY EN DESACUERDO	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	7	100	7	100	7	100	7	100	7	100	7	100
MEDIA DEL INDICADOR	7,14											
CATEGORÍA DEL INDICADOR	MODERADO NIVEL											
MEDIA DE LA DIMENSIÓN	22,06											
CATEGORÍA DE LA DIMENSIÓN	MODERADO NIVEL											

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Gráfico 5. Frecuencia Relativa sobre los Factores Institucionales

En relación a lo observado en el gráfico N° 5, se muestra para el ítem 7 el 57,14% de la directiva manifestó que de acuerdo que los profesores es personal calificado y posee experiencia sobre el Programa de Ingeniería en Sistemas,

asimismo, un 42,85% que de acuerdo y un 0% que ni de acuerdo, ni en desacuerdo, en desacuerdo y muy en desacuerdo, lo cual indica que la directiva se encuentra segura del personal calificado que posee el Programa de Ingeniería en Sistemas.

Asimismo, para el ítem 8, el 57,14% de la directiva manifestó que de acuerdo que el Programa de Ingeniería en Sistemas brinda apoyo psicológico a los estudiantes, asimismo, un 42,85% que de acuerdo y un 0% que ni de acuerdo, ni en desacuerdo, en desacuerdo y muy en desacuerdo, lo cual muestra que la directiva está segura que el programa brinda apoyo psicológico a la población estudiantil.

Por otro lado para el ítem 9, se evidenció que el 57,14% de la directiva manifestó que de acuerdo que el Programa de Ingeniería en Sistemas presenta flexibilidad curricular, asimismo, un 42,85% que de acuerdo y un 0% que ni de acuerdo, ni en desacuerdo, en desacuerdo y muy en desacuerdo, lo cual pauta que la directiva asegura que el programa posee flexibilidad curricular a la población estudiantil.

Para el ítem 10, se muestra que el 42,85% de la directiva respondieron que de acuerdo el Programa de Ingeniería en Sistemas ofrece consultas pedagógicas a los docentes, de igual manera, un 28,57% muy de acuerdo y ni de acuerdo, ni en desacuerdo y un 0% en desacuerdo y muy en desacuerdo, lo cual indica que la directiva asegura que su personal docente le ofrecen consultas pedagógicas.

Seguidamente, el ítem 11, mostró que el 42,85% de la directiva respondió que ni de acuerdo, ni en desacuerdo que el Programa de Ingeniería en Sistemas entrega Becas y formas de financiamiento a los estudiantes, asimismo, un 28,57% de muy de acuerdo y en desacuerdo y un 0% que de acuerdo y muy en desacuerdo, lo cual muestra que la directiva en ciertas oportunidades expresa que el programa otorga becas o formas de financiamiento a los estudiantes.

Por último, para el ítem 12, se evidencia que el 42,85% de la directiva mostró que de acuerdo y en desacuerdo el Programa de Ingeniería en Sistemas apoya a los estudiantes mediante convenios de pago, asimismo, un 14,28% de muy de acuerdo y un 0% ni de acuerdo, ni en desacuerdo y muy en desacuerdo, lo cual conduce a que la directiva en varias oportunidades apoya a la población estudiantil con becas o convenios de pago.

Estos resultados, muestran controversias a lo planteado por Báez y Pedraza, (2011), quienes señalan que los factores institucionales han reconocido que los mismos comprueban un alto porcentaje en la determinación de la deserción estudiantil, lo cual modela cuestiones a lo visto en el Sistema de Información Académico (SIA) recopilados de los procesos de Admisión en la Universidad de la Guajira Sede Maicao, demostrando que el mismo tiene escasos factores institucionales en el Programa de Ingeniería en Sistemas.

Asimismo, lo pautado por Yorke y Longden (2004), quienes mencionan que en este campo de factores institucionales se destacan la calidad de los programas y los procesos de formación, la flexibilidad curricular, las practicas pedagógicas de los docentes, becas y formas de financiamiento, orden público, entorno político, nivel de interacción personal, apoyo académico y psicológico y todo aquello que estimule la vinculación del estudiante con el medio académico e investigativo.

Variable: Deserción estudiantil

Dimensión: Factores determinantes

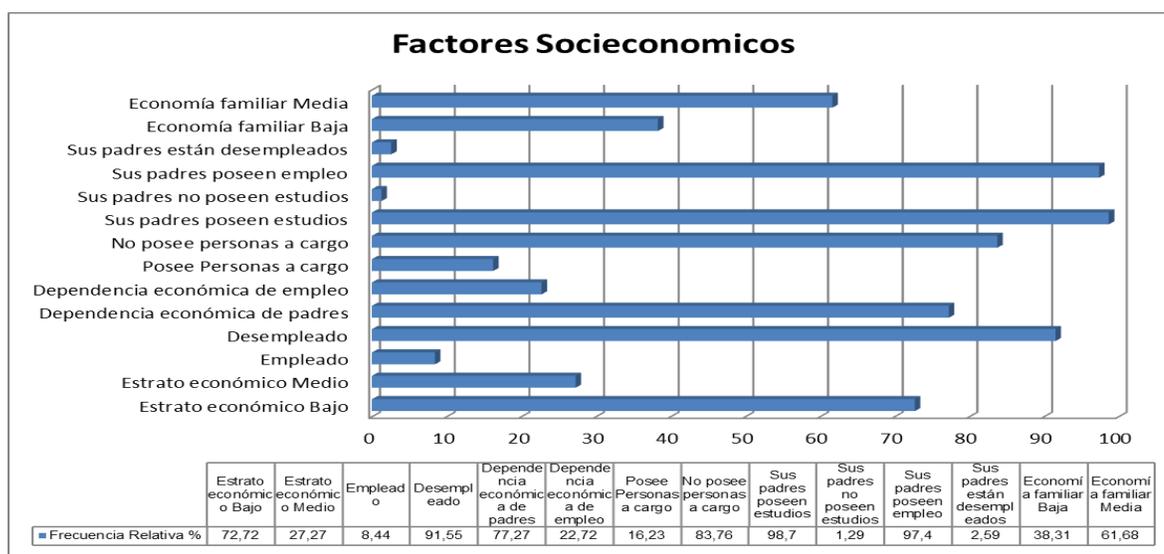
Indicador: Factores socioeconómicos

Grupo de Muestro: Estudiantes

Tabla 6. Factores Socioeconómicos

INDICADOR	FACTORES SOCIECONOMICOS													
	Ítem 15		Ítem 16		Ítem 17		Ítem 18		Ítem 19		Ítem 20		Ítem 21	
ALTERNATIVAS DE RESPUESTAS	Estrato económico Bajo	Estrato económico Medio	Empleado	Desempleado	Dependencia económica de padres	Dependencia económica de empleo	Posee Personas a cargo	No posee personas a cargo	Sus padres poseen estudios	Sus padres no poseen estudios	Sus padres poseen empleo	Sus padres están desempleados	Economía familiar Baja	Economía familiar Media
Frecuencia Relativa %	72,72	27,27	8,44	91,55	77,27	22,72	16,23	83,76	98,70	1,29	97,40	2,59	38,31	61,68
TOTAL	100		100		100		100		100		100		100	
MEDIA DEL INDICADOR	49,99													
CATEGORÍA DEL INDICADOR	MODERADO NIVEL													
DESVIACION ESTANDAR DEL INDICADOR	36,96													
CATEGORIA DEL INDICADOR	MODERADO NIVEL													

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Grafico 6. Frecuencia Relativa sobre los Factores Socioeconómicos

Según los datos mostrados en el gráfico N° 6, se evidencia que en ítem 15 nivel de estrato el 72,72% de los estudiantes posee un estrato bajo, además, en el ítem 16 situación laboral, un 91,55% de los estudiantes se encuentran desempleados,

en el ítem 17 dependencia económica, se observó que el 77,27% de los estudiantes dependen económicamente de sus padres, en el ítem 18 personas a cargo un 83,76% de los estudiantes no poseen personas a cargo, en el ítem 18 estudios de los padres, se mostró que el 98,7% de los padres poseen estudios.

Asimismo, en el ítem 19 situación laboral de los padres, se evidencia que el 97,4% de los padres poseen empleo, en el ítem 20 situación económica familiar se reveló que un 61,68% poseen una economía media, todos estos resultados muy poco afectan la deserción estudiantil debido a que la desviación estándar de este indicador factores socioeconómicos fue de 36,95 clasificándola como una alta desviación lo cual se interpreta como nula dependencia de los factores socioeconómicos con la variable deserción estudiantil.

Estos resultados presentan controversias a lo planteado por Báez y Pedraza (2011), quienes muestran que el asunto financiero es determinante en la posibilidad de vincularse y mantenerse en el sistema educativo. Si las instituciones y el sistema en general no cuentan con políticas claras de financiación y con las sinergias de interacción necesarias entre las diferentes dependencias será muy difícil de lograr el acceso y la permanencia de los estudiantes.

Asimismo, acceder, mantenerse y graduarse requiere inversión básicamente en matrícula, transporte, libros y alimentación. Los hogares de estratos 1 y 2 deben comprometer su patrimonio para tal efecto; es fundamental la estabilidad de los ingresos, si estos caen entonces los rubros educativos disminuyen para mantener otras necesidades básicas.

Identificar los recursos necesarios para la aplicación de la minería de datos en la Universidad de la Guajira.

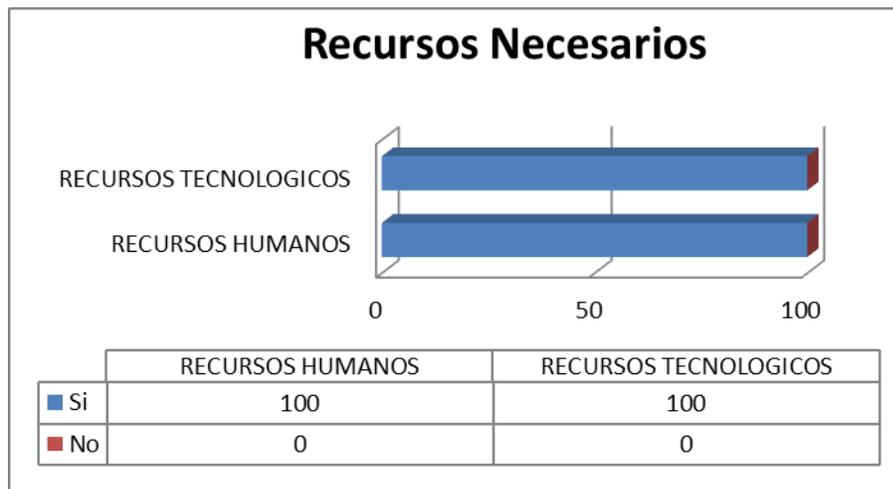
Variable: Minería de datos

Dimensión: Recursos Necesarios

Tabla 7. Recursos Necesarios

DIMENSIÓN	RECURSOS NECESARIOS			
INDICADORES	RECURSOS HUMANOS		RECURSOS TECNOLOGICOS	
	Si	No	Si	No
FRECUENCIA RELATIVA	Fr %	Fr %	Fr %	Fr %
Ítem 1	100	0		
Ítem 2	100	0		
Ítem 3			100	0
Ítem 4			100	0
MEDIA DEL INDICADOR	100		100	
CATEGORÍA DEL INDICADOR	Alto Nivel		Alto Nivel	
MEDIA DE LA DIMENSIÓN	100			
CATEGORÍA DE LA DIMENSIÓN	Alto Nivel			

Fuente: Elaboración propia, (2019).



Fuente: Elaboración propia, (2019).

Grafico 7. Frecuencia Relativa sobre los Recursos Necesarios

En relación a los datos mostrados en el gráfico N° 7, se evidencia que el 100% de los recursos humanos y tecnológicos se encuentran en el programa de ingeniería en sistemas, es decir, el laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira cuenta con un personal, a su vez, el personal capacitado para ejecutar la minería de datos, como también, existe el Hardware y el Software necesario para ejecutar la minería de datos en el laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira, lo cual indica que dicha

facultad de ingeniería posee tanto el recurso humano como el tecnología para efectuar la minería de datos.

Estos resultados son justificados por la Enciclopedia de Ejemplos (2017), la cual señala que el Hardware y Software se muestra de la siguiente manera: En computación, los términos hardware y software aluden a aspectos distintos de todo sistema informático: el aspecto físico y el digital respectivamente, cuerpo y alma de toda computadora. El hardware es el conjunto de las partes físicas que integran el cuerpo de un sistema computarizado: las placas, circuitos, mecanismos y dispositivos eléctricos, así como de procesamiento, soporte y conexión.

Además, a lo señalado por Lledo, (2011), quien afirma que los recursos humanos tienen un enfoque de aplicación y práctica de las actividades más importantes dentro de la organización o empresas siendo la gestión del talento humano un pilar fundamental para el desarrollo exitoso de los procesos, pues al final las personas son los responsables de ejecutarlas actividades porque los proyectos no se desarrollan por si solos. Dentro de los recursos humanos que se utilizan en la minería de datos se encuentran los ingenieros en sistemas, los técnicos en informática y analistas de datos

Describir el modelo predictivo en minería de datos que se ajuste a la dinámica de la deserción estudiantil.

Variable: Minería de datos

Dimensión: Modelos predictivos

En el presente trabajo de investigación se pretende determinar las causas de deserción del programa con datos proporcionados por el estudiante, mediante un cuestionario y entrevista estructurada. Para ello se utilizó una técnica de minería de datos conocida como Árbol de decisión, el cual entra dentro del Modelo predictivo de retención, la cual es una técnica de clasificación y predicción, donde

se clasifican los casos en grupos con características propias y se pronostica una variable dependiente o criterio que es la deserción en el programa.

Con el Árbol de decisión se puede determinar el perfil del estudiante que va a incurrir en deserción y además ¿Cuáles son las principales variables del cuestionario que influyen en la deserción?.

Para la creación del Árbol de decisión se utilizó el método CHAID Exhaustivo, del programa SPSS, versión 23. La variable dependiente o a predecir fue la deserción en el programa y del cuestionario fueron seleccionadas 11 variables explicativas o independientes, cada una con dos categorías o modalidades, seleccionadas de acuerdo a la revisión de trabajos anteriores y del criterio del investigador relacionado con las posibles variables que influyen en la deserción del programa. Estas variables fueron: Edad, Sexo, Estado Civil, Posición entre hermanos, Apego familiar, Apoyo económico de los padres, Apoyo educativo de los padres, Estrato Socioeconómico, Tenencia de empleo y la satisfacción con el programa.

Antes de crear el Árbol de decisión, se procedió a realizar un análisis de tablas de contingencia o cruzadas entre cada una de las variables independientes y la dependiente, usando un prueba Chi-cuadrado y el estadístico coeficiente de contingencia, para observar la significancia de la asociación entre las dos variables.

Luego de realizadas todas las pruebas, se obtuvo que la única variable independiente con significancia estadística fue la satisfacción con el programa ($X^2 = 5,500$; $P = .019$). Estos resultados se observan en las tablas 9 y 10. Esto se hace generalmente para seleccionar variables influyentes en la dependiente y contrastar con los resultados del árbol de decisión.

Tabla 8. Recuentos de la tabla cruzada Piensa desertar del programa*Satisfacción con el programa

Tabla cruzada. Recuento						
				Satisfacción con el programa		Total
				No	Si	
Piensa desertar del programa	No	40	39	79		
	Si	24	51	75		
Total		64	90	154		

Fuente: Programa SPSS versión 23.

Tabla 9. Valores de los estadísticos y significancia de las pruebas

Pruebas de Chi-cuadrado					
	Valor	Gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	5,500 ^a	1	,019		
Corrección de continuidad ^b	4,759	1	,029		
Razón de verosimilitud	5,544	1	,019		
Prueba exacta de Fisher				,022	,014
Asociación lineal por lineal	5,464	1	,019		
N de casos válidos	154				

Fuente: Programa SPSS versión 23.

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 31,17.

b. Sólo se ha calculado para una tabla 2x2.

Tabla 10. Valor y significancia del Coeficiente de Contingencia

Medidas simétricas			
		Valor	Significación aproximada
Nominal por Nominal	Coeficiente de contingencia	,186	,019
N de casos válidos		154	

Fuente: Programa SPSS versión 23.

Resultados de la aplicación del Árbol de Decisión

En la tabla 11 de resumen del Árbol de Decisión, se presentan las especificaciones y resultados resumidos. En este sentido, se observa que se aplicó el método CHAID Exhaustivo, la variable dependiente e independientes, no se realizó una validación del modelo, la máxima profundidad del árbol fue de 3, con casos mínimos en nodo padre de 100 e hijo de 50. La variable independiente que influyó en la deserción fue la satisfacción con el programa, aparecieron solo 3 nodos o cajas, 2 nodos terminales en la parte baja, con una profundidad de 1.

Tabla 11. Resumen del modelo de Árbol de Decisión obtenido

RESUMEN DEL MODELO		
Especificaciones	Método de crecimiento	CHAID Exhaustivo
	Variable dependiente	Piensa desertar del programa
	Variables independientes	Edad en años, Sexo, Estado Civil, Posición entre hermanos, Apego familiar, Satisfacción con el programa, Apoyo económico de los padres, Apoyo educativo de los padres, Estrato Socioeconómico, Tiene empleo
	Validación	Ninguna
	Máxima profundidad del árbol	3
	Casos mínimos en nodo padre	100
	Casos mínimos en nodo hijo	50
Resultados	Variables independientes incluidas	Satisfacción con el programa
	Número de nodos	3
	Número de nodos terminales	2
	Profundidad	1

Fuente: Programa SPSS versión 23.

Relacionado con la clasificación en forma correcta del modelo se observa en la tabla 11, que el modelo de retención bajo la técnica usada de Árbol de Decisión clasifica en forma correcta globalmente el 59,1% de los casos y para cada categoría de la variable dependiente se observa un acierto mucho mayor en la categoría de deserción o que va a dejar el programa (68,0%) comparado con la categoría de que no va a dejar el programa donde el porcentaje es del 50,6%. Hay que tener en cuenta que en esta investigación la muestra resultó muy pequeña para la utilización del modelo de retención mediante la clasificación de la técnica de minería de datos del Árbol de decisión, generalmente se utilizan más de 500 casos.

Tabla 12. Porcentajes de clasificación correcta

Observado	Clasificación		
	Pronosticado		
	No	Si	Porcentaje correcto
No	40	39	50,6%
Si	24	51	68,0%
Porcentaje global	41,6%	58,4%	59,1%

Fuente: Programa SPSS versión 23

Método de crecimiento: CHAID Exhaustivo

Variable dependiente: Piensa desertar del programa o alta satisfacción del programa.

CONCLUSIONES

En relación a los resultados mostrados se presentan las siguientes conclusiones: en el primer objetivo el cual se debe a Caracterizar los factores determinantes de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira, se mostró que los resultados sobre los factores individuales no afectan la deserción estudiantil debido a que la desviación estándar de este indicador factores individuales fue de 35,05 clasificándola como una alta desviación lo cual se interpreta como nula dependencia de los factores individuales con la variable deserción estudiantil.

Además, en los factores académicos, se evidenció un nivel moderado, que representa variaciones como existencia de un gran número de estudiantes que utilizan la carrera de trampolín para cambiarse a otra carrera, desajustes en el programa, debido a que existe un numero considerado que no le dedica el tiempo necesario al programa, desequilibrio en el programa debido a que existe numerosa población estudiantil que muy poco posee buena relación con los docentes, desajuste dado a que la mayoría de los estudiantes no se encuentra seguro que las exigencias del programa causan o no deserción en el mismo, causando un inquietud por parte de los docentes y directiva del programa.

A su vez, que menos de la mitad de los estudiantes lo consideran y otros no, lo que genera desequilibrio en la implementación o modificación del contenido programático, desequilibrio en el nivel de satisfacción que tienen algunos estudiantes por el programa, y por ende busca a la directiva indagar las causas del porque el alumnado no posee satisfacción del programa, un nivel moderado, que representa desviaciones como presencia de un gran número de estudiantes que piensa desertar del programa, generando falencias al programa por la poca población inscrita.

Por otro lado, se evidencia que los docentes muy poco poseen conocimiento si los estudiantes conocen todo lo relacionado al programa de ingeniería en sistemas, que muy poco se interesan si el estudiante ingresa al programa por vocación, no muestran interés en conocer las razones por las cuales el estudiante ingresa al programa de ingeniería en sistema, asimismo, conocen en algunas oportunidades que los estudiantes ingresan o no al programa por la falta económica, que poco conocen si los estudiantes presentan inconvenientes para asistir al horario de clases del programa, que conocen que la universidad le ofrece las rutas estudiantiles a su población, que varias oportunidades los docentes evidencia gracias a lo físico y las actitudes que el estudiante piensa en desertar.

Además, los docentes en ciertas oportunidades han evidenciado que los aspectos económicos influyen a que el estudiante deserte del programa de ingeniería en sistemas, conocen los pocos hábitos de estudios que demuestran los estudiantes antes y después de clases, que muy poco conocen si los estudiantes poseen compromiso al programa de ingeniería en sistemas y que los estudiantes en la mayoría de los casos muestran motivación de sus clases.

Consecuentemente, la directiva posee conocimiento que el programa posee un alto porcentaje de deserción estudiantil, que en ciertas ocasiones observa que el alumnado posee conocimientos sobre el perfil del programa, que en la mayoría de los casos poco conoce si los estudiantes dedican las horas suficientes al estudio del contenido programático del programa de ingeniería en sistemas, que en la mayoría están al tanto que el programa de ingeniería en sistemas posee una alta calidad, que en ciertas oportunidades enuncia que el programa en ingeniería en sistemas ejecuta una prueba de admisión para el ingreso al mismo y conoce poco sobre la base de datos que posee el programa de ingeniería en sistemas.

En concordancia con los factores institucionales, se evidenció que la directiva se encuentra segura del personal calificado que posee el Programa de Ingeniería en Sistemas, que está segura que el programa brinda apoyo psicológico a la

población estudiantil, que asegura que el programa posee flexibilidad curricular a la población estudiantil, que asegura que su personal docente le ofrecen consultas pedagógicas, que en ciertas oportunidades expresa que el programa otorga becas o formas de financiamiento a los estudiantes y que en varias oportunidades apoya a la población estudiantil con becas o convenios de pago.

Por último, en consideración a los factores socioeconómicos se mostró que muy poco afectan la deserción estudiantil debido a que la desviación estándar de este indicador factores socioeconómicos fue de 36,95 clasificándola como una alta desviación lo cual se interpreta como nula dependencia de los factores socioeconómicos con la variable deserción estudiantil.

Estos resultados presentan controversias a lo planteado por El Ministerio Nacional de educación (MEN). (2014), quien desarrolla un resumen del estado del arte de los factores determinantes de la deserción agrupando las variables más utilizadas en cuatro categorías: individuales, académicos, institucionales y socioeconómicos. Igualmente, afirma que existe consenso en que la deserción estudiantil es el resultado del efecto no de una sola categoría, sino del efecto individual y de la interacción de diferentes categorías de factores.

Sin embargo, en términos generales y de acuerdo con la revisión de la literatura, se puede decir que existen más trabajos que destacan la perspectiva institucional y en los que los diferentes conjuntos de variables (institucionales, socioeconómicas, académicas y personales) son analizados de manera independiente y no como un conjunto de factores que determinan la decisión de desertar.

Por otro lado, en relación al segundo objetivo, el cual es Identificar los recursos necesarios para la aplicación de la minería de datos en la Universidad de la Guajira, se evidencia que los recursos humanos y tecnológicos se encuentran en el programa de ingeniería en sistemas, es decir, el laboratorio de la facultad de

ingeniería en sistema de la Universidad de la Guajira cuenta con un personal, a su vez, el personal capacitado para ejecutar la minería de datos, como también, existe el Hardware y el Software necesario para ejecutar la minería de datos en el laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira, lo cual indica que dicha facultad de ingeniería posee tanto el recurso humano como el tecnología para efectuar la minería de datos.

Sin embargo, estos resultados presentan similitudes a lo mostrado por la Enciclopedia de Ejemplos (2017), muestra el Hardware y Software de la siguiente manera: En computación, los términos hardware y software aluden a aspectos distintos de todo sistema informático: el aspecto físico y el digital respectivamente, cuerpo y alma de toda computadora. El hardware es el conjunto de las partes físicas que integran el cuerpo de un sistema computarizado: las placas, circuitos, mecanismos y dispositivos eléctricos, así como de procesamiento, soporte y conexión.

Por otro lado, Lledo, (2011), afirma que los recursos humanos tienen un enfoque de aplicación y práctica de las actividades más importantes dentro de la organización o empresas siendo la gestión del talento humano un pilar fundamental para el desarrollo exitoso de los procesos, pues al final las personas son los responsables de ejecutarlas actividades porque los proyectos no se desarrollan por si solos.

Consecuentemente, mediante el tercer objetivo el cual es Describir el modelo predictivo en minería de datos que se ajuste a la dinámica de la deserción estudiantil, se evidenció que la clasificación en forma correcta del modelo se observa que en la técnica usada Árbol de Decisión clasifica en forma correcta globalmente el 59,1% de los casos y para cada categoría de la variable dependiente se observa un acierto mucho mayor en la categoría de deserción o que va a dejar el programa (68,0%) comparado con la categoría de que no va a dejar el programa donde el porcentaje es del 50,6%. Hay que tener en cuenta que

en esta investigación la muestra resultó muy pequeña para la utilización de la técnica de clasificación como el Árbol de decisión del modelo predictivo de retención, generalmente se utilizan más de 500 casos.

Estos resultados muestran similitudes por lo planteado por Vincent Tinto (2012), quien menciona que el modelo de retención que tal vez haya tenido la mayor aceptación, además que ha vinculado variadas teorías y de diferentes exponentes, y propone que el éxito de un estudiante está mediado por su grado de integración académica y social; es decir, por el cumplimiento de estándares académicos y por los lazos que se tejan entre él, su entorno y los agentes de la institución.

Asimismo, Tinto propone un cambio de enfoque sobre el fenómeno de persistencia y un modelo de acción institucional. En este contexto, Tinto provee una distinción conceptual entre los términos de retención y persistencia. La persistencia la define desde la perspectiva del estudiante, quien tiene el control de permanecer o no en la institución y en la carrera elegida hasta graduarse.

El autor reflexiona en torno a diversas condiciones mínimas que deben considerar las universidades para crear estas comunidades, entre las cuales destaca el esfuerzo sistemático de esclarecer lo que se necesita para tener éxito en el proceso de formación. La idea esencial es proveer al estudiante información de calidad que le ayude a razonar por los diferentes cursos, a lo largo de la carrera y en su estancia en la universidad.

A la final, mediante le último objetivo el cual es Proponer mediante la minería de datos un modelo para la predicción de la deserción estudiantil en el programa de Ingeniería en Sistemas de la Universidad de la Guajira, se evidenció que el programa de Ingeniería en Sistemas de la Universidad de la Guajira debe diseñar un modelo de minería de datos para la predicción de la deserción estudiantil, como también que la institución implemente estrategias para que el estudiante se

desarrolle cognitiva y afectivamente en su progreso académico hasta la obtención del título universitario.

En tal sentido, debido a que la deserción estudiantil en el sistema de educación superior es un problema que afecta a todos y es sinónimo de exclusión social, se torna urgente buscar las soluciones que aplaquen de mejor manera las causas de este fenómeno que afecta, no solo a la promoción de las personas, sino también, al desarrollo general de la sociedad.

RECOMENDACIONES

En concordancia con los resultados mostrados se muestran las siguientes recomendaciones:

- ✚ El personal docente y directivo necesitan integrarse a lo que desea el estudiante en relación al programa de ingeniería en sistemas.
- ✚ Extraer datos de diferentes fuentes para aplicar las técnicas de minería.
- ✚ Aplicar los instrumentos necesarios para recabar información y poder diseñar modelos predictivos más asertivos que determinen el nivel de deserción estudiantil.
- ✚ Diseñar las diferentes técnicas de minería de datos en el laboratorio del programa de ingeniería en sistemas.
- ✚ Mantener actualizado el recurso humano en el laboratorio del programa de ingeniería en sistemas.
- ✚ Se recomienda un licenciamiento en cuanto al software SPSS, para el tratamiento de los datos.
- ✚ Se recomienda aplicar la técnica de minería de datos de árbol de decisión en el laboratorio de ingeniería en sistemas un mayor número de casos es decir con la data de toda la población estudiantil de la Universidad de la Guajira.

REFERENCIAS BIBLIOGRAFICAS

Abensur, S. (2009). *Factores socioeconómicos y personales relacionados con la deserción estudiantil en la Escuela de Negocios Internacionales de la Facultad de Ciencias Económicas y Negocios de la Universidad Nacional de la Amazonía Peruana, 2002-2006*. (Tesis de Grado no Publicada). Perú: Universidad Nacional Mayor de San Marcos. Obtenido de:

http://www.cybertesis.edu.pe/sisbib/2009/abensur_ds/pdf/abensur_ds.pdf.

Acevedo, D.; Torres, J. y Tirado, D. (2015). *Análisis de la deserción estudiantil en el programa de ingeniería de alimentos de la Universidad de Cartagena durante el periodo académico 2009-2013*, *Formación Universitaria*, 8 (1), 35-42 (2015).

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Dept. of Education Office of Educational Research and Improvement.

Anuies. (2007). *Retención y deserción en un grupo de Instituciones Mexicanas de Educación Superior*. Biblioteca de la educación superior. México.

Amador P. (2018). *Técnico en Informática*. Obtenido de:
<https://ude.edu.uy/fci-ude/tecnico-en-informatica/>

Amaya, Y.; Barrientos, E.; Heredia, D. (2016). *Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos*. Universidad Francisco de Paula Santander, Ocaña, Colombia. Barranquilla, Colombia.

Arranz, J.; Cid, J. y Muro, J. (2000). *La duración del desempleo en la Argentina*. *Anales de la Asociación Argentina de Economía Política*. Buenos Aires: AAEP.

Báez, C. y Pedraza, P. (2011). *Persistencia y graduación: Hacia un modelo de retención estudiantil para instituciones de educación superior*. Bogotá: Universidad de la Sabana – Colciencias.

Balestrini M. (2006). *Como se elabora el proyecto de investigación*. Caracas: BL Consultores Asociados. 7ma edición.

Barragán, S.; Moreno, L.; González, T. (2017). *Acercamiento a la deserción estudiantil desde la integración social y académica*. Artículo. Revista de la Educación Superior Vol. 46.

Behar D. (2008). *Metodología de la Investigación*. Ediciones Shalom.
file:///C:/Users/casa2/Documents/Libro%20metodologia%20investigacion.DANIEL BEHAR.(2008).pdf

Bensimon, E. M. et al. (2004). *Doing research that makes a difference*. Journal of Higher, Ohio, v. 1, p. 104-126.

Bernal C. A. *Metodología de la Investigación*. 2006. México. Editorial Pearson.

Bitecna. (2018). *Hardware y Software*. Obtenido de:
https://unctad.org/divs/gds/dmfas/fr/who/Documents/DMFAS_6_HardwareSoftware_6_SP.pdf

Boado, M. (2005). *Una aproximación a la deserción estudiantil universitaria en Uruguay*. Universidad de la República, Montevideo, Uruguay, en cooperación con el Instituto Internacional para la Educación Superior en América Latina y el Caribe, pp. 10–24.

Braxton, J.; Milem, J. y Shaw, A. (2000). *The influence of Active Learning on the College Student Departure Process: Toward a Revision of Tinto's Theory*. The Journal of Higher Education, 71 (5), 569-590. Obtenido de:

DOI: <http://dx.doi.org/10.2307/2649260>

Cabrera, A. et al. (2012). *Pathways to a four-year degree: determinants of transfer and degree completion*. In: SEIDMAN, Alan (Ed.). *College student retention: a formula for student Success*. Westport: Praeger, p. 155-209.

Cabrera, L.; Bethencourt, J.; González, M. y Álvarez, P. (2006). *Un estudio transversal retrospectivo sobre la prolongación y abandono de estudios universitarios*. *Revista Electrónica de Investigación y Evaluación Educativa*, 12(1), 105-127.

Carmona, E. (2014). *Tutorial sobre Máquinas de Vectores Soporte*. (SVM). Obtenido de:

[http://www.ia.uned.es/~ejcarmona/publicaciones/\[2013Carmona\]%20SVM.pdf](http://www.ia.uned.es/~ejcarmona/publicaciones/[2013Carmona]%20SVM.pdf)

Castaño, E., Gallón, S., Gómez, K. y Vásquez, J. (2007). *Deserción estudiantil universitaria: una aplicación de modelos de duración*. *Lecturas de economía*, 60, 39-66.

Castejón, O. (2011). *Diseño y análisis de experimentos con Statistix*. Fondo Editorial Biblioteca Universidad Rafael Urdaneta. Obtenido de:

<http://www.uru.edu/fondoeditorial/libros/pdf/manualdestatistix/occompleto.pdf>

Castillo, M.; Osorio, A. y Montero, S. (2010). *Deserción y retención, en la carrera de Economía de la Pontificia Universidad Javeriana Cali: un análisis de supervivencia, 2000-2008*. *Economía, Gestión y Desarrollo*, 9, 11- 33.

Cerezo, R.; Bernardo, A.; Núñez, J.C.; Tuero, E. y Esteban, M. (2015). *Predicción del abandono universitario: variables Explicativas y medidas de prevención*. Artículo. Revista Fuentes, 16. Universidad de Oviedo.

Chiavenato, I. (2009). *Gestión del talento humano Tercera edición*. México. McGraw-Hill/Interamericana Editores, S.A. de C.V.

CIDSE. (2011). *Análisis de la herramienta SPADIES diseñada por el Ministerio de Educación Nacional y el CEDE*. Recuperado. Obtenido de:
http://www.alfaguia.org/alfaguia/files/1320349109_33.pdf

Díaz, C. (2008). *Modelo Conceptual Para La Deserción Estudiantil Universitaria Chilena*. Estudios Pedagógicos XXXIV, N° 2: 65-86

Domènech J. (1992). *Una aplicación del análisis de la supervivencia en ciencias de la salud*. Anuario de Psicología, 55, 109-141.

Donoso, S. y Schiefelbein, E. (2007). *Análisis de los modelos explicativos de retención de estudiantes en la universidad: una visión desde la desigualdad social*. Estudios Pedagógicos XXXIII, N°1, pp: 7-27, 2007.

Ecured. (2019). Deserción escolar. Obtenido de:
www.ecured.cu/desercion_escolar

Educaweb. (2018). *Técnico en Informática*. Obtenido de:
<https://www.educaweb.com/profesion/tecnico-informatico-servicios-soporte-408/>

Emagister. Guía de Orientación. (2017). *Ingeniero en Sistemas*. Obtenido de:
<https://www.emagister.com/que-hace-un-ingeniero-de-sistemas/>

Enciclopedia de Ejemplos. (2017). *Hardware y Software*. Obtenido de:

<https://www.ejemplos.co/30-ejemplos-de-hardware-y-software/>

Ethington, C. (1970). *A psychological model of student persistence*. *Research in Higher education*. 31, (31): Pp 279 – 293.

Etxebarria, M. (2017). *Gestión de recursos humanos*. Obtenido de:
<http://atalayagestioncultural.es/capitulo/gestion-recursos-humanos>

Fidias A. (2012). *Metodología de la investigación científica*. 7ma Edición. Editorial Episteme. Caracas. pp. 23.

Finol M. y Camacho H. (2008). *El proceso de investigación científica*. Segunda edición. McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V. México.

Fisc. (2018). *Técnico en Informática para la Gestión Empresarial*. Universidad Tecnológica de Panamá. Obtenido de:
<http://www.fisc.utp.ac.pa/tecnico-en-informatica-para-la-gestion-empresarial>

Fozda R, B., Kumar, L. & Kannan, S. (2006). *Study of the Factors Responsible for the Dropout from the BSc Programme of Indira Gandhi National Open University*. *International Review of Research in Open and Distance Learning*, 3, 1492-3831.

Gaviria, A. (2002). *Los que suben y los que bajan: educación y movilidad sociales en Colombia*, en Fedesarrollo, Santafé de Bogotá, Alfaomega.

Gill, H.; Rao, P.; Hatos, E. (2001). *Data warehousing: la integración de información para la mejor toma de decisiones*. México: Prentice-Hall Hispanoamericana.

Giovagnoli, P. (2002). *Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración*. Documento de Trabajo.

Departamento de economía, facultad de ciencias económicas, Universidad de la Plata.

González, L. (2005). *Estudio sobre la repitencia y deserción en la educación superior chilena*. IESALC – UNESCO. Obtenido de: www.iesalc.unesco.org.ve

Guzmán, C.; Durán, D.; Franco, J.; Castaño, E.; Gallón, S.; Gómez, K. y Vásquez, J. (2009). *Deserción estudiantil en la educación superior colombiana*, Bogotá. Colombia: Imprenta Nacional de Colombia.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques: concepts and techniques*. Elsevier.

Heredia, D. y Nieto, W. (2011). *Generación de un modelo predictivo para determinar el desempeño académico en la asignatura fundamentos de programación II del programa de Ingeniería de Sistemas*. Colombia.

Hernández R.; Fernández C. y Baptista P. (2014). *Metodología de la investigación*. (6ta ed.). México: McGraw-Hill.

IBM Corp. Released (2015). *IBM SPSS Statistics for Windows*, version 23:0. Armonk, NY: IBM Corp.

lingen.unam. (2018). *Ingeniería de sistemas*. Obtenido de: <http://www.iingen.unam.mx/esmx/Investigacion/Especialidad/Paginas/IngenieriaDeSistemas.aspx> universidad nacional autónoma de México.

Ishitani, T. (2006). *Studying attrition and degree completion behavior among firstgeneration college students in the United States*. *The Journal of Higher Education*, 77(5), 861-885.

Kelley-Winstead D. (2013). *New Directions in Education Research: Using Data Mining Techniques to Explore Predictors of Grade Retention*.

Kumar, S.; Bharadwaj, B. y Pal, S. (2012). *Mining Education Data to Predict Student's Retention: A comparative Study*. India.

Lledo, P. (2011). *Administrador de Proyectos. El ABC para un director de proyectos exitoso*. Victoria, BC, Canadá. 3ra edición.

Lopera, C. (2008). *Determinantes de la deserción universitaria en la Universidad del Rosario*. Borradores de investigación, 95.

Manyika, J.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Hung, B. (2011). *Big data: The next frontier for innovation, competition, and prod.*

MBA & Educación Ejecutiva. (2018). *Gestión de carrera. Cómo debe ser un buen analista de datos: el trabajo del futuro*. Obtenido de:

<https://mba.americaeconomia.com/articulos/reportajes/como-debe-ser-un-buen-analista-de-datos-el-trabajo-del-futuro>.

Matheus, A.; Claudio, R.; Marcelo, T.; Gutiérrez, L.; Reyes, G. (2018). *Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins Santiago de Chile*, Artículo en la Revista Educ. Pesqui., São Paulo, Vol. 44. Obtenido de:

<http://www.scielo.br/pdf/ep/v44/1517-9702-ep-44-e172094.pdf>

MEN. (2017). *¿Qué es el SPADIES? - Sistemas información*. Obtenido en:

<http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-254648.html>

Méndez, C. (2009). *Metodología. Diseño y desarrollo del proceso de investigación con énfasis en ciencias empresariales*. 4ta edición. México. Editorial Limusa.

Merlino, A.; Herrero, V.; Ayllón, S.; Escanés, G. (2013). *Aplicación de un modelo de duración en programas de prevención de deserción universitaria REDIE*. Artículo. Revista Electrónica de Investigación Educativa, vol. 15, núm. Universidad Autónoma de Baja California Ensenada, México.

Ministerio de Educación Nacional República de Colombia (2019). Deserción escolar. Obtenido de: www.mineducacion.gov.co

Ministerio de Educación Nacional. (2014). *Informe Estadísticas de Educación Superior*. Obtenido de:

http://www.mineducación.gov.co/sistemasdeinformacion/1735/articles212350_Estadisticas_de_Educación_Superior_.pdf

Ministerio de educación nacional (2012). *Informe SPADIES*. Obtenido de:

http://spadies.mineducación.gov.co/spadies/consultas_predefinidas.html?2

Ministerio de educación nacional (2009). *Deserción Estudiantil en la Educación Superior Colombiana: Metodología de Seguimiento, Diagnóstico y Elementos para su prevención*. Bogotá: Ministerio de Educación Nacional.

Montero, R. (2007). *¿Cuánto dura el desempleo de la población más pobre en Chile?*. Cuadernos de Economía, 44, 211-231.

Morales, E. (2009). *Descubrimiento de conocimiento en bases de datos*. Material en línea. Obtenido de:

<http://ccc.inaoep.mx/~emorales/Cursos/KDD/principal.html>

Patiño L.; Cardona A. (2012). *Review of some studies on university student dropout in Colombia and Latin America*. Acta Universitaria, 23 (4). 37 – 46.

Pérez, J. y Merino, M. (2009). *Definicion.de: Definición de ingeniería de sistemas*.
Obtenido de: (<https://definicion.de/ingenieria-de-sistemas/>)

Provost, F., & Fawcett, T. (2013). *Data Science for Business. What you need to know about data mining and data-analytic thinking* (First Edition).

Quinlan J. R. (1986). *Induction of Decision Trees*.

Rodríguez, C; Sotomonte, J.; Montenegro, C.; Gaona, P y Castellanos, J. (2016). *Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos*. Artículo. Revista Científica/ No. 26/ BOGOTÁ, D.C.

Rodríguez, J.; Rojas, E.; Franco, R. (2017). *Clasificación de datos usando el método K-NN*. Revista Vínculos, Universidad Distrital. Volumen 4 número 1 pagina 10. Obtenido de:
<http://revistas.udistrital.edu.co/ojs/index.php/vinculos/article/viewFile/4111/5778>

Rodríguez, J. y Hernández, M. (2008). *La deserción escolar universitaria en México. La experiencia de la Universidad Autónoma Metropolitana*. Actualidades Investigativas en Educación, 8 (1).

Rodríguez, V.; Campos, J.; Aguilera, J. (2013). *Modelo predictivo para la permanencia en la educación superior Línea Temática: Factores asociados. Tipos y perfiles de abandono*. Universidad de Playa Ancha de Ciencias de la Educación – Valparaíso, Chile. Obtenido de:
<http://revistas.utp.ac.pa/index.php/clabes/article/view/1588>

Salas, M. (2007). *The transition from higher education to employment in Europe: the analysis of the time to obtain the first job*. Higher Education, 54(3), 333-360.

Saldaña, M. y Barriga, O. (2010). *Adaptación de un modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile*. Revista de ciencias sociales, 16 (4), 616-628. Departamento Nacional de Planeación Pública. (2014). Colombia.

Sen, A. (2000). *Desarrollo y libertad*, Barcelona, Ediciones Planeta, Ediciones Paidós Ibérica.

Spady, W. G. (1970a). *Dropouts from Higher Education: An Interdisciplinary Review and Synthesis*. Interchange, 1(1), 64–85.

Tamayo y Tamayo, M. (2007). *El Proceso de la Investigación Científica*. 3ª edición. Limusa. México DF. Noriega Editores.

Timarán, R.; Calderón, A. y Jiménez J. (2013). *Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil*. Revista Ventana Informática Nº 28. Universidad de Manizales. Colombia.

Timaran, R. y Jimenez J. (2014). *Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM*. Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación. Artículo 758.

Tinto, V., & Cullen, J. (1975). *Dropout from Higher Education: A Theoretical Synthesis of Recent Research*. Review of Educational Research, 45(1), 89–125.

Tinto, V. (1985). *Limits of Theory and Practice in Student Attrition*. The Journal of Higher Education, 53(6), 687.

Tinto, V. (2012). *Taking student retention seriously*. Syracuse University.

UNESCO (2004). *Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura*. Repetition at high cost in Latin America and the Caribbean. IESALC - UNESCO. Obtenido de:
http://www.unesco.org/fileadmin/MULTIMEDIA/HQ/ED/ED_new/pdf/LAC-GEM-2014ENG.pdf

Universidad de Murcia (2014). *Hardware y Software*.
<https://www.um.es/docencia/barzana/VARIOS/Recursos-hardware-software.html>

Universia. (2018). *¿Qué puedo estudiar para ser analista de datos?*. Obtenido de:
<http://noticias.universia.com.ar/educacion/noticia/2018/03/01/1158218/puedo-estudiar-analista-datos.html>

Valero, S. (2009). *Desarrollo de una herramienta de análisis de datos para predecir deserción escolar en la Universidad Tecnológica de Izúcar de Matamoros*. Tesis no publicada. Universidad Tecnológica de Izúcar de Matamoros. México.

Valero, O. (2009). *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos*.

Vercellis, C. (2009). *Business intelligence: data mining and optimization for decision making*. Editorial John Wiley and Sons.

ANEXOS

ANEXO A.



MINERÍA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE LA GUAJIRA

Autor:

Diego Madrid

INSTRUMENTO A

**UNIVERSIDAD DE LA GUAJIRA
FACULTAD: INGENIERÍA EN SISTEMAS
MAICAO - 2019**

INSTRUCCIONES

En el presente cuestionario usted encontrará una serie de preguntas ante las cuales deberá hacer su elección. Asegúrense de leer detenidamente cada ítem antes de contestarlo. Cada pregunta presenta dos o más alternativas, usted deberá escoger la que más identifique.

Marque con una (X) la casilla correspondiente de acuerdo a lo anteriormente señalado. Se le agradece responder con suma objetividad y no dejar ninguna pregunta sin respuesta. De ser necesario exprese las dudas con el investigador.

Instrumento para aplicar a los Estudiantes del Programa de Ingeniería en Sistema.

INDICADOR: FACTORES INDIVIDUALES	SELECCIÓN DE ALTERNATIVA				
ESTUDIANTES	MARQUE CON UNA X				
1. Edad.	a. Menores de 18 años <input type="radio"/>				
	b. Entre 18 y 30 años <input type="radio"/>				
	c. Entre 30 y 45 años. <input type="radio"/>				
2. Genero.	a. Masculino <input type="radio"/>				
	b. Femenino <input type="radio"/>				
	c. Otro <input type="radio"/>				
3. Estado Civil.	a. Soltero <input type="radio"/>				
	b. Casado <input type="radio"/>				
	c. Divorciado <input type="radio"/>				
	d. Viudo <input type="radio"/>				
4. Posición entre los hermanos.	a. Mayor <input type="radio"/>				
	b. Entre mayor y Menor <input type="radio"/>				
	c. Menor <input type="radio"/>				
5. Señale una o varias de estas opciones que prevalecen en su entorno familiar.	a. Apego familiar <input type="radio"/>				
	b. Valores consolidados <input type="radio"/>				
	c. Maltrato Psicológico <input type="radio"/>				
	d. Maltrato Físico <input type="radio"/>				
	e. Embarazo Precoz <input type="radio"/>				
6. Entre usted y su entorno familiar, como es su Salud.	a. Sana <input type="radio"/>				
	b. Enfermedad <input type="radio"/>				
7. Señala cual es el apoyo de sus padres en su formación.	a. Económico <input type="radio"/>				
	b. Educativo <input type="radio"/>				
INDICADOR: FACTORES ACADÉMICOS	ALTERNATIVA DE RESPUESTA				
ESTUDIANTES	MD	D	ND/NED	ED	MED
8. Ingresó al Programa de Ingeniería en Sistemas por efecto de obtener un cupo universitario para luego cambiarse de carrera.					
9. Dedicó tiempo suficiente a la programación e investigación del Programa de Ingeniería en Sistemas.					

10. Existe buena relación con los docentes del programa.					
11. Las exigencias del Programa de Ingeniería en Sistemas causan deserción.					
12. El contenido programático es suficiente para el Programa de Ingeniería en Sistemas.					
13. Posee un alto nivel de satisfacción sobre el Programa de ingeniería en sistemas.					
14. Piensa desertar al Programa de Ingeniería en Sistemas.					
INDICADOR: FACTORES SOCIOECONÓMICOS	SELECCIÓN DE ALTERNATIVAS				
ESTUDIANTES					
15.Cuál es su Estrato Económico	a. Alto	<input type="radio"/>			
	b. Medio	<input type="radio"/>			
	c. Bajo	<input type="radio"/>			
16. Cuál es su situación laboral	a. Trabajo independiente	<input type="radio"/>			
	b. Empleado	<input type="radio"/>			
	c. No Trabaja	<input type="radio"/>			
17. Depende económicamente	a. Padres	<input type="radio"/>			
	b. Empleo	<input type="radio"/>			
	c. Familiares	<input type="radio"/>			
18. Posee personas a cargo	a. Si	<input type="radio"/>			
	b. No	<input type="radio"/>			
19. Cuál es el nivel de estudio de sus padres	a. Sin estudios	<input type="radio"/>			
	b. Primaria	<input type="radio"/>			
	c. Secundaria	<input type="radio"/>			
	d. Universitaria	<input type="radio"/>			
20. Cuál es la situación laboral de su familia.	a. Trabajo independiente	<input type="radio"/>			
	b. Empleado	<input type="radio"/>			
	c. No Trabaja	<input type="radio"/>			
21. Como es la economía en su entorno familiar	a. Alta	<input type="radio"/>			
	b. Media	<input type="radio"/>			
	c. Baja	<input type="radio"/>			

ANEXO B



MINERÍA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE LA GUAJIRA

Autor:

Diego Madrid

INSTRUMENTO B

**UNIVERSIDAD DE LA GUAJIRA
FACULTAD: INGENIERÍA EN SISTEMAS
MAICAO - 2019**

INSTRUCCIONES

En el presente cuestionario usted encontrará una serie de preguntas ante las cuales deberá hacer su elección. Asegúrense de leer detenidamente cada ítem antes de contestarlo. Cada afirmación presenta (5) alternativas, usted deberá escoger la que más identifique, las alternativas de respuestas son las siguientes:

MD: Muy de Acuerdo

D: De Acuerdo

ND/NE: Ni de Acuerdo, ni en Desacuerdo

ED: En Desacuerdo

MED: Muy en Desacuerdo

Marque con una (X) la casilla correspondiente a la escala de acuerdo a lo anteriormente señalado. Se le agradece responder con suma objetividad y no dejar ninguna pregunta sin respuesta. De ser necesario exprese las dudas con el investigador.

Instrumento para aplicar a los Docentes del Programa de Ingeniería en Sistema.

INDICADOR:FACTORES ACADEMICOS	MD	D	ND/NE	ED	MED
1. Los estudiantes poseen conocimientos previos sobre el programa de ingeniería en sistemas.					
2. El ingreso de los estudiantes al programa es por vocación.					
3. El ingreso del estudiante al programa es por falta de oportunidades.					
4. El ingreso del estudiante al programa es por falta de recursos económicos.					
5. Los estudiantes del Programa presentan inconvenientes para asistir a clases en el horario establecido.					
6. La universidad beneficia a los estudiantes con el servicio de rutas estudiantiles urbanas e intra – urbanas.					
7. Identifica por medio de observaciones físicas y mentales cuales son estudiantes desertores.					
8. Los aspectos económicos influyen en la deserción estudiantil del programa de ingeniería en sistemas.					
9. Los estudiantes manifiestan poco hábitos de estudios en los momentos antes y después del acto de clase.					
10. Existe compromiso educativo al programa por parte de los estudiantes.					
11. Los estudiantes del programa de ingeniería en sistemas presentan motivación en la clase.					

ANEXO C



MINERÍA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE LA GUAJIRA

Autor:

Diego Madrid

INSTRUMENTO C

**UNIVERSIDAD DE LA GUAJIRA
FACULTAD: INGENIERÍA EN SISTEMAS
MAICAO - 2019**

INSTRUCCIONES

En el presente cuestionario usted encontrará una serie de preguntas ante las cuales deberá hacer su elección. Asegúrense de leer detenidamente cada ítem antes de contestarlo. Cada afirmación presenta (5) alternativas, usted deberá escoger la que más identifique, las alternativas de respuestas son las siguientes:

MD: Muy de Acuerdo

D: De Acuerdo

ND/NE: Ni de Acuerdo, ni en Desacuerdo

ED: En Desacuerdo

MED: Muy en Desacuerdo

Marque con una (X) la casilla correspondiente a la escala de acuerdo a lo anteriormente señalado. Se le agradece responder con suma objetividad y no dejar ninguna pregunta sin respuesta. De ser necesario exprese las dudas con el investigador.

Instrumento para aplicar a la Directiva del Programa de Ingeniería en Sistema.

INDICADOR: FACTORES ACADÉMICOS	ALTERNATIVA DE RESPUESTA				
DIRECTIVA	MD	D	ND/NE	ED	MED
1. El Programa de Ingeniería en Sistemas posee un alto índice de estudiantes reprobados.					
2. El alumnado conoce el perfil de ingreso del programa de ingeniería en sistemas.					
3. Los estudiantes dedican horas suficientes al estudio aun cuando conocen los niveles de exigencia de la carrera.					
4. El Programa de Ingeniería en Sistemas posee una alta Calidad.					
5. El programa de Ingeniería en Sistemas ejecuta una prueba de admisión para el ingreso al mismo.					
6. En el Programa de Ingeniería en Sistemas posee una base de datos de estudiantes que ingresan, desertan y egresan.					
INDICADOR: FACTORES INSTITUCIONALES	ALTERNATIVA DE RESPUESTA				
DIRECTIVA	MD	D	ND/NE	ED	MED
7. Los profesores es personal calificativo y posee experiencia sobre el Programa de Ingeniería en Sistemas.					
8. En el Programa de Ingeniería en Sistemas brinda apoyo psicológico a los estudiantes.					
9. El Programa de Ingeniería en Sistemas presenta flexibilidad curricular.					
10. El Programa de Ingeniería en Sistemas ofrece consultas pedagógicas a los docentes.					
11. El Programa de Ingeniería en Sistemas entrega Becas y formas de financiamiento a los estudiantes.					
12. El Programa de Ingeniería en Sistemas apoya a los estudiantes mediante convenios de pago.					

ANEXO D



MINERÍA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN EL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE LA GUAJIRA

Autor:

Diego Madrid

**INSTRUMENTO D
LISTA DE COTEJO**

**UNIVERSIDAD DE LA GUAJIRA
FACULTAD: INGENIERÍA EN SISTEMAS
MAICAO - 2019**

INSTRUCCIONES

En el presente cuestionario usted encontrará una serie de preguntas ante las cuales deberá hacer su elección. Asegúrense de leer detenidamente cada ítem antes de contestarlo. Cada afirmación presenta 2 alternativas, para seleccionar una, usted deberá escoger la que más identifique, las alternativas de respuestas son las siguientes:

Si

No

Marque con una (X) la casilla correspondiente a lo anteriormente señalado. Se le agradece responder con suma objetividad y no dejar ninguna pregunta sin respuesta. De ser necesario exprese las dudas con el investigador.

Lista de Cotejo para el Laboratorio de la Facultad de Ingeniería en Sistemas de la Universidad de la Guajira

INDICADOR: RECURSOS HUMANOS	ALTERNATIVA DE RESPUESTA	
DIRECTIVA	SI	NO
1. El laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira cuenta con un personal.		
2. En el laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira existe personal capacitado para ejecutar la minería de datos.		
INDICADOR: RECURSOS TECNOLOGICOS	ALTERNATIVA DE RESPUESTA	
DIRECTIVA	SI	NO
3. Existe el Hardware necesario para ejecutar la minería de datos en el laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira.		
4. Existe el Software necesario para ejecutar la minería de datos en el laboratorio de la facultad de ingeniería en sistema de la Universidad de la Guajira.		