



UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS

CARRERA DE INGENIERÍA EN INFORMÁTICA Y SISTEMAS COMPUTACIONALES

PROYECTO DE INVESTIGACIÓN

MODELO BASADO EN MINERÍA DE DATOS PARA DETERMINAR FACTORES DE DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI.

Proyecto de Titulación presentado previo a la obtención del Título de Ingeniería en Informática y
Sistemas Computacionales

AUTORAS:

Masabanda Yépez Jhenny Flor

Zapata Rocha Carla Jhoana

TUTORA:

Ing. MSc. Albán Taipe Mayra Susana

LATACUNGA – ECUADOR

Agosto 2019



DECLARACIÓN DE AUTORÍA

Nosotras **Masabanda Yépez Jhenny Flor**, con cédula de ciudadanía No. - **050354752-3** y **Zapata Rocha Carla Jhoana**, con cédula de ciudadanía No. - **050453900-8**, declaramos ser autoras del presente proyecto de investigación: **“MODELO BASADO EN MINERÍA DE DATOS PARA DETERMINAR FACTORES DE DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI”**, siendo la Ing. MSc. **Mayra Susana Albán Taipe**, con cédula de ciudadanía No. **050231198-8** tutora del presente trabajo; y eximio expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Jhenny Flor Masabanda Yépez
CC: 050354752-3

Carla Jhoana Zapata Rocha
CC: 050453900-8

AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN

En calidad de Tutora del Trabajo de Investigación sobre el título: **“MODELO BASADO EN MINERÍA DE DATOS PARA DETERMINAR FACTORES DE DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI”**, de las señoritas estudiantes, **Masabanda Yépez Jhenny Flor**, con cédula de ciudadanía No.- **050354752-3** y **Zapata Rocha Carla Jhoana**, con cédula de ciudadanía No.- **050453900-8**, de la carrera Ingeniería en Informática y Sistemas Computacionales, considero que dicho Informe Investigativo cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometidos a la evaluación del Tribunal de Validación de Proyecto que el Honorable Consejo Académico de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi designe, para su correspondiente estudio y calificación.

Latacunga, Julio, 2019



Tutora de Titulación

Nombre: Ing. MSc. Mayra Susana Albán Taipe
CC: 050231198-8



APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueban el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotopaxi, y por la **FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS**; por cuanto, el o los postulantes: **Masabanda Yépez Jhenny Flor**, con cédula de ciudadanía No.- **050354752-3** y **Zapata Rocha Carla Jhoana**, con cédula de ciudadanía No.- **050453900-8**, con el título de Proyecto de titulación: **MODELO BASADO EN MINERÍA DE DATOS PARA DETERMINAR FACTORES DE DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI**, han considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido al acto de Sustentación de Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, Julio, 2019

Para constancia firman:

Lector 1 (Presidente)

Nombre: Cadena Moreano José Augusto
CC: 050155279-8

Lector 2

Nombre: Bravo Mullo Silvia Jeaneth
CC: 050243712-2

Lector 3

Nombre: Llano Casa Alex Christian
CC: 050258986-4

AGRADECIMIENTO

A Dios, por haberme dado la vida, por tener una familia maravillosa y la oportunidad de conocer a personas de lindos sentimientos de las cuales he aprendido a ser perseverante, a no dejarme vencer por los obstáculos.

También agradezco a mis maestros por haberme formado profesionalmente en especial a la Ing. MSc. Mayra Albán, en calidad de tutora por la paciencia otorgada en guiarme e impartir sus conocimientos.

Jhenny

AGRADECIMIENTO

A mis amigos con todos los que compartí dentro y fuera de las aulas, que se convierten en amigos de vida y aquellos que serán mis colegas, gracias por todo su apoyo y diversión.

A toda mi familia que han estado presente cada una de las etapas de mi vida.

Carla

DEDICATORIA

El presente proyecto les dedico a mis padres María y Carlos y a mis hermanos Adrián y Vanessa, quienes han estado conmigo durante toda mi vida, brindándome su amor y apoyo incondicional, enseñándome que con esfuerzo, dedicación y entrega total que puede alcanzar las metas que nos proponemos en la vida.

A Mauricio quien me ha brindado su amor y palabras de aliento para continuar con mis estudios y a mi hija Greta por las veces que me ha hecho sonreír con su inocencia.

Jhenny

DEDICATORIA

Dedico este trabajo principalmente a Dios, por haberme dado la vida y permitirme haber llegado hasta este momento tan importante de mi formación profesional.

A mis padres Gladys y Jorge por ser el pilar más importante, por demostrarme siempre su cariño, amor y apoyo incondicional sin importar nuestras diferencias de opiniones y a mi hermanita Maite.

Carla

ÍNDICE GENERAL

DECLARACIÓN DE AUTORÍA	ii
AVAL DEL TUTOR DEL PROYECTO DE TITULACIÓN	iii
APROBACIÓN DEL TRIBUNAL DE TITULACIÓN	iv
AGRADECIMIENTO.....	v
DEDICATORIA	vii
ÍNDICE GENERAL.....	ix
ÍNDICE DE TABLAS	xii
ÍNDICE DE FIGURAS.....	xiii
RESUMEN.....	xvi
ABSTRACT.....	xvii
AVAL DE TRADUCCIÓN	xviii
1 INFORMACIÓN GENERAL	1
2 JUSTIFICACIÓN DEL PROYECTO.....	2
3 BENEFICIARIOS DEL PROYECTO	3
4 EL PROBLEMA DE INVESTIGACIÓN.....	3
4.1 Definición del problema.....	5
5 OBJETIVOS:	5
5.1 General	5
5.2 Específicos	5
6 ACTIVIDADES Y SISTEMA DE TAREAS EN RELACIÓN A LOS OBJETIVOS PLANTEADOS.....	6
7 FUNDAMENTACIÓN CIENTÍFICO TÉCNICA.....	7
7.1 Antecedentes	7
7.2 Principales referentes teóricos.....	7
7.3 Bases teóricas	9
7.3.1 Deserción Universitaria.....	9
7.3.2 Modelo para Predecir	9

7.3.3	Tipos de deserción.....	13
7.3.4	Factores de deserción	14
7.3.5	Minería de datos	15
7.3.6	Las técnicas de minería de datos (Data Mining)	15
7.3.7	Campos de aplicación de la minería de datos en el Ecuador.....	18
7.3.8	Ventajas y desventajas de la minería de datos.....	19
7.3.9	Las herramientas de minería de datos	19
7.3.10	Metodología de minería de datos	21
7.3.11	Técnicas de Preprocesado	24
7.3.12	Métricas para la evaluación de clasificadores	25
7.3.13	DataSet	27
7.3.14	Excel.....	27
8	HIPÓTESIS	28
9	METODOLOGÍAS Y DISEÑO EXPERIMENTAL	28
9.1	Metodología	28
9.1.1	Metodología Científica.....	28
9.1.2	Tipos de investigación.....	28
9.1.3	Métodos de investigación.....	29
9.1.4	Técnicas de investigación.....	29
9.1.5	Instrumentos de investigación.....	30
9.1.6	Métodos específicos	31
9.2	Diseño de la Investigación	34
10	ANÁLISIS Y DISCUSIÓN DE RESULTADOS	35
10.1	Recopilación de información	35
10.2	Confiabilidad de los datos	37
10.3	Análisis de la información.....	37
10.4	Etapas de especificación del modelo	39
10.4.2	Análisis de datos a través de técnicas de minería de datos.....	54

11	IMPACTOS (INSTITUCIONALES, SOCIALES, AMBIENTALES O ECONÓMICOS)	75
11.1	Impacto institucional	75
11.2	Impacto económico	75
11.3	Impacto social	75
12	PRESUPUESTO PARA LA PROPUESTA DEL PROYECTO.....	76
12.1	Gastos directos	76
12.2	Gastos indirectos	76
12.3	Gastos aproximados	76
13	CONCLUSIONES Y RECOMENDACIONES	77
13.1	Conclusiones	77
13.2	Recomendaciones.....	77
14	BIBLIOGRAFÍA.....	79
15	ANEXOS.....	87
15.1	Anexo 1: Hoja de vida del grupo de trabajo.....	87
15.2	Anexo 2: Encuesta.....	90
15.3	Anexo 4: Análisis de la encuesta.....	100

ÍNDICE DE TABLAS

Tabla 1: Actividades de los objetivos	6
Tabla 2: Lista de variables planteadas por Bean 1980	12
Tabla 3: Cuadro comparativo de las metodologías de minería de datos	23
Tabla 4: Descripción de Variables	28
Tabla 5: Carreras de la Facultad de Ciencias de la Ingeniería y Aplicadas	30
Tabla 6: Resultados de la evaluación de la escala de Likert	36
Tabla 7: Resultados de validez de la encuesta	37
Tabla 8: Estadísticas descriptivas de Ciencias de Ingeniería y Aplicadas	37
Tabla 9: Variables (x, y).....	41
Tabla 10: Descripción de los factores de deserción estudiantil.....	43
Tabla 11: Estimación del modelo inicial.....	44
Tabla 12: Modelo con todas variables	44
Tabla 13: Modelo ajustado de factores de deserción estudiantil.....	46
Tabla 14: Los resultados del proceso de la regresión lineal.....	49
Tabla 15: Método: BestFirst - Atributo: CfsSubsetEval	55
Tabla 16: Método: GreedyStepwise - Atributo: CfsSubsetEval	56
Tabla 17: Factores de deserción.....	56
Tabla 18: Especificaciones del Multilayer Perceptrón.....	61
Tabla 19: Resultado de la herramienta que utiliza Weka que es Multilayer Perceptron.....	62
Tabla 20: Especificaciones del Navie Bayes.....	63
Tabla 21: Resultado de la herramienta que utiliza Weka que es Naive Bayes.....	64
Tabla 22: Resultado de la herramienta que utiliza Weka que es J48	65
Tabla 23: Especificaciones del J48	67
Tabla 24: Resultado de la herramienta que utiliza Weka que es Random Forest	68
Tabla 25: Especificaciones del Random Forest	69
Tabla 26: Resultado de la herramienta que utiliza Weka que es SMO	70
Tabla 27: Especificaciones del SMO	71
Tabla 28: Instancias clasificadas por técnicas de minería de datos.....	72
Tabla 29: Resultados de las técnicas de minería de datos que posee Weka.....	72
Tabla 30: Descripción gastos directos.....	76
Tabla 31: Descripción gastos indirectos	76
Tabla 32: Total, Gastos Directos e Indirectos	76

ÍNDICE DE FIGURAS

Figura 1: Modelo de Fishbein y Ajzen, 1975.....	9
Figura 2: Modelo de Ethington, 1990.....	10
Figura 3: El modelo de Spady, 1970.....	11
Figura 4: Modelo de Tinto, 1987, Adaptado de Tinto (1975).....	11
Figura 5: Modelo de Bean, 1985.....	13
Figura 6: Modelo de Pascarella y Terenzini	13
Figura 7: Pasos que componen la metodología KDD	23
Figura 8: Familias técnicas en preparación de datos.....	24
Figura 9: Familias técnicas en reducción de datos.....	25
Figura 10: Diseño de la investigación para determinar factores de deserción.....	34
Figura 11: Modelo conceptual	40
Figura 12: Ecuación generada.....	42
Figura 13: Residuos del modelo con la variable 16.....	47
Figura 14: Continuidad del residuo del modelo.....	47
Figura 15: Probabilidad Jarque-Bera	48
Figura 16: Regresión lineal.....	51
Figura 17: Estructura gráfica de un modelo Naive Bayes	59
Figura 18: Árbol de decisión.....	65
Figura 19: ThresholdCurve con 1 - Naive Bayes.....	73
Figura 20: ThresholdCurve con 1 - J48	73
Figura 21: ThresholdCurve con 1 - SMO	73
Figura 22: ThresholdCurve con 1 – RandomForest.....	73
Figura 23: ThresholdCurve con 1 – Multilayer Perceptrón	74
Figura 24: CostCurve con 1 - Naive Bayes.....	74
Figura 25: CostCurve con 1 - J48	74
Figura 26: CostCurve con 1 – SMO	74
Figura 27: CostCurve con 1 – RandomForest.....	74
Figura 28: CostCurve con 1 – Multilayer Perceptrón	75
Figura 29: Género de los estudiantes	100
Figura 30: Figura 27: Edad de los encuestados.....	100
Figura 31: Carreras de Ciencias de la Ingeniería y Aplicadas	101
Figura 32: Ciclos al que pertenece los estudiantes	102
Figura 33: Estado civil de los estudiantes.....	102
Figura 34: Etnia de los estudiantes	103
Figura 35: Posee discapacidad	104

Figura 36: Tipo de discapacidad de posee	104
Figura 37: Grado de discapacidad que posee.....	105
Figura 38: Porcentaje de discapacidad que posee.....	105
Figura 39: Porcentaje de discapacidad que posee.....	106
Figura 40: País de origen de los estudiantes	106
Figura 41: Lugar de procedencia de los estudiantes de Eléctrica	107
Figura 42: Lugar de procedencia de los estudiantes de Industrial	108
Figura 43: Lugar de procedencia de los estudiantes de Electromecánica.....	109
Figura 44: Lugar de procedencia de los estudiantes de Electromecánica.....	110
Figura 45: Lugar de residencia actual de los estudiantes de Eléctrica.....	111
Figura 46: Lugar de residencia actual de los estudiantes de Industrial.....	111
Figura 47: Lugar de residencia actual de los estudiantes de Electromecánica.....	112
Figura 48: Lugar de residencia de los estudiantes de Sistemas de Información	113
Figura 49: Estatus social de los estudiantes	113
Figura 50: Vive cerca de la universidad.....	114
Figura 51: Tipo de hogar de los estudiantes	115
Figura 52: Número total del miembro de la familia del estudiante.....	115
Figura 53: Tipo de vivienda en la que reside	116
Figura 54: Ingresos familiares mensuales.....	116
Figura 55: Origen de recursos de los estudios	117
Figura 56: Horas que se dedican a trabajar	118
Figura 57: Nivel de formación del padre	118
Figura 58: Nivel de formación de la madre	119
Figura 59: Colegio del que proceden	120
Figura 60: Elección de carrera	120
Figura 61: Proceso de formación académica	121
Figura 62: Proceso de nivelación.....	122
Figura 63: Expectativas de la carrera.....	122
Figura 64: Expectativas de la carrera.....	123
Figura 65: Comportamiento negativo de docentes en el aula de clases.....	124
Figura 66: Comportamiento negativo de docentes en el aula de clases.....	124
Figura 67: Razones de conductas inapropiadas	125
Figura 68: No tener amistades	126
Figura 69: Espacios de bienestar estudiantil adecuados	126
Figura 70: Bullying por parte de compañeros.....	127
Figura 71: Bullying por parte de docentes	128

Figura 72: Expectativas profesionales planeadas.....	128
Figura 73: Abandonar la universidad.....	129
Figura 74: Razones para dejar la universidad	129
Figura 75: Rendimiento académico de los estudiantes	130
Figura 76: Razones de bajo rendimiento académico	131
Figura 77: Valoran el esfuerzo los docentes	131
Figura 78: Inadecuado sistema de tutorías.....	132
Figura 79: Problemas familiares	133
Figura 80: Inadecuada formación del docente	133
Figura 81: Inadecuada orientación del aprendizaje.....	134
Figura 82: Inadecuado desarrollo de cualidades humanas del docente.....	135
Figura 83: Inadecuado desarrollo de cualidades humanas del docente.....	135
Figura 84: Estado de gestación del estudiante planificado - no planificado	136
Figura 85: Estado de gestación del estudiante planificado - no planificado	137
Figura 86: Factor bullying	138
Figura 87: Factor bullying	138
Figura 88: Actividades que se dedica el estudiante	139
Figura 89: Actividades que se dedica el estudiante	140
Figura 90: Factores que impiden continuar sus estudios	141
Figura 91: Factores relacionados con la etapa académica	142
Figura 92: Estado de ánimo frente al docente.....	143

UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS

TÍTULO: “MODELO BASADO EN MINERÍA DE DATOS PARA DETERMINAR FACTORES DE DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI”.

Autoras: Masabanda Yépez Jhenny Flor
Zapata Rocha Carla Jhoana

RESUMEN

La deserción en las universidades es considerada como el abandono de los estudiantes que puede ser temporal o definitivo, causa efectos negativos como problemas socioeconómicos para el estudiante y la institución que lo acogió. Por tal razón, este problema debe ser tratado a profundidad para establecer estrategias que permitan minimizar las tasas de deserción en las universidades y la culminación exitosa de los estudios universitarios. La revisión de la literatura permitió determinar estudios para predecir la deserción a través de técnicas de minería de datos, sin embargo, los estudios analizados fueron ejecutados en entornos de Educación Superior que difieren del contexto de educación que se aplica en el Ecuador. Se propone un modelo para determinar factores de deserción estudiantil a través de técnicas de minería de datos para determinar factores que influyen en la deserción y su influencia predictiva. El proceso experimental se basa en una encuesta en línea aplicada a 1457 estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas de las Carreras de Ingenierías: Eléctrica, Sistemas de Información, Electromecánica e Industrial. La metodología aplicada corresponde a Knowledge Discovery in Databases (KDD), que consta de cinco etapas: selección, pre-procesamiento, transformación, extracción e interpretación y evaluación. Los resultados encontrados permiten determinar que los factores: conducta en el aula de clases, bullying, motivación del docente – alumno, limitado conocimiento de la asignatura, adicción de las redes sociales, estado emocional, conocimiento adquirido en los cursos de nivelación, formación académica, sistema de ingreso a la universidad, problemas familiares, residencia y expectativas respecto a la carrera seleccionada, son los factores que tienen mayor influencia en la deserción de los estudiantes en la Facultad de Ciencias de la Ingeniería y Aplicadas. Mientras que las técnicas de minería de datos J48, Random Forest y Sequential Minimal Optimization (SMO), dieron como resultado una tasa de predicción de la deserción del 92%. Se concluye que el uso de técnicas de minería de datos puede ser consideradas como importantes para realizar estudios de las causales que afectan a los estudiantes en su permanencia estudiantil universitaria. Además, esta herramienta podría ser considerada como una herramienta de apoyo para las autoridades universitarias a fin de que se establezca estrategias y políticas que permitan mitigar las tasas de deserción.

Palabras claves: deserción universitaria, factores de deserción, minería de datos, Machine Learning.

TECHNICAL UNIVERSITY OF COTOPAXI
FACULTY OF ENGINEERING AND APPLIED SCIENCES

TITLE: "MODEL BASED ON DATA MINING TO DETERMINE STUDENT DROPOUT FACTORS IN THE FACULTY OF SCIENCES OF THE ENGINEERING AND APPLIED OF THE TECHNICAL UNIVERSITY OF COTOPAXI"

Authors: Masabanda Yépez Jhenny Flor
Zapata Rocha Carla Jhoana

ABSTRACT

Attrition in universities is considered as the abandonment of students that may be temporary or permanent, causing negative effects such as socioeconomic problems for the student and the institution that hosted him. For this reason, this problem must be treated in depth to establish strategies that minimize the drop-out rates in universities and the successful completion of university studies. The literature review allowed to determine studies to predict attrition through data mining techniques, however, the studies analyzed were executed in Higher Education environments that differ from the context of education that is applied in Ecuador. A model is proposed to determine student dropout factors through data mining techniques to determine factors that influence dropout and its predictive influence. The experimental process is based on an online survey applied to 1457 students of the Faculty of Engineering and Applied Sciences of Engineering Careers: Electrical, Information Systems, Electromechanical and Industrial. The methodology applied corresponds to Knowledge Discovery in Databases (KDD), which consists of five stages: selection, pre-processing, transformation, extraction and interpretation and evaluation. The results found allow determining that the factors: inappropriate behavior in the classroom, bullying, teacher-student motivation, limited knowledge of the subject, addiction of social networks, emotional state, knowledge acquired in leveling, academic training, system of University entrance, family problems, residence and expectations regarding the selected career, are the factors that have the greatest influence on the drop-out of students in the Faculty of Engineering and Applied Sciences. While the J48, Random Forest and Sequential Minimal Optimization (SMO) data mining techniques resulted in a 92% dropout prediction rate. It is concluded that the use of data mining techniques can be considered as important for studies of the causes that affect students during their university student stay. In addition, this tool could be considered as a support tool for university authorities in order to establish strategies and policies to mitigate dropout rates.

Keywords: university desertion, desertion factors, data mining, Machine Learning.



Universidad
Técnica de
Cotopaxi

CENTRO DE IDIOMAS

AVAL DE TRADUCCIÓN

En calidad de Docente del Idioma Inglés del Centro de Idiomas de la Universidad Técnica de Cotopaxi; en forma legal **CERTIFICO** que: La traducción del resumen del proyecto de investigación al Idioma Inglés presentado por las señoritas Egresadas de la Carrera de **INGENIERÍA EN INFORMÁTICA Y SISTEMAS COMPUTACIONALES** de la **FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS, MASABANDA YÉPEZ JHENNY FLOR Y ZAPATA ROCHA CARLA JHOANA**, cuyo título versa “**MODELO BASADO EN MINERÍA DE DATOS PARA DETERMINAR FACTORES DE DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI**”, lo realizaron bajo mi supervisión y cumple con una correcta estructura gramatical del Idioma.

Es todo cuanto puedo certificar en honor a la verdad y autorizo a las peticionarias hacer uso del presente certificado de la manera ética que estimaren conveniente.

Latacunga, Julio del 2019.

Atentamente,

Lic. José Ignacio Andrade.
C.C. 0503101040
DOCENTE CENTRO DE IDIOMAS



1 INFORMACIÓN GENERAL

Título del Proyecto:

Modelo basado en minería de datos para determinar factores de deserción estudiantil en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

Fecha de inicio:

Octubre 2018.

Fecha de finalización:

Agosto 2019.

Lugar de ejecución:

Barrio El Ejido, Parroquia Eloy Alfaro, Cantón Latacunga, Provincia de Cotopaxi, Zona 3, Facultad de Ciencias de la Ingeniería y Aplicadas - Universidad Técnica de Cotopaxi.

Facultad que auspicia:

Ciencias de la Ingeniería y Aplicadas.

Carrera que auspicia:

Ingeniería en Informática y Sistemas Computacionales.

Proyecto de investigación vinculado:

Modelamiento de Algoritmos para Sistemas de Información.

Equipo de Trabajo:

Tutora de Titulación

Ing. MSc. Mayra Susana Albán Taípe.

Coordinadores del Proyecto

Jhenny Flor Masabanda Yépez.

Carla Jhoana Zapata Rocha.

Área de Conocimiento:

Área: Ciencias.

Sub – Área: Informática.

Línea de investigación:

Tecnologías de la Información y Comunicación (TIC's) y Diseño Gráfico.

Sub líneas de investigación de la Carrera:

Inteligencia Artificial e Inteligencia de Negocios.

2 JUSTIFICACIÓN DEL PROYECTO

La deserción estudiantil universitaria, es un problema complejo que conlleva impactos negativos para las universidades Latinoamericanas, entre ellos sobresalen los problemas socioeconómicos, sociales, académicas y personales del alumno. En el contexto de educación en Chile, tres de cada diez estudiantes desertan de las instituciones universitarias en el primer año. Los datos precedentes de Chile permiten deducir sobre las consecuencias de la deserción, como el impacto sobre el endeudamiento familiar y el perjuicio que conlleva en los primeros quintiles de ingreso (Pérez et al. 2018).

Por tal razón, las instituciones de educación superior han empezado a trabajar en el diseño de mecanismos y formas de asegurar el aumento de la retención en los primeros años de estudios universitarios. Muchos estudios aplicados para predecir la deserción siguen como patrón el modelo teórico planteado por (Vicent Tinto, 1975) en donde el autor manifiesta sobre los indicadores de retención y persistencia, haciendo referencia a la retención como el compromiso de la universidad y la persistencia en el esfuerzo del estudiante. De la misma manera en el modelo de (William Spady, 1970), incorporan factores externos a los percibidos por el individuo, con un enfoque sociológico, el autor considera factores relacionados al medio familiar. Los factores planteados por (John Bean, 1985), corresponde a factores académicos, psicosociales, ambientales y de socialización (Ramírez and Grandón, 2018).

Por otro lado, el modelo predictivo empleado por (Cuji, Gavilanez & Sánchez 2017) y citado por (Ramírez & Grandón 2018) busca predecir futuros comportamientos (Ramírez & Grandón, 2018), por medio de la técnica de clasificación que corresponde a árboles de decisión empleando el J48. El autor realizó la predicción a través del uso de variables como: el semestre que cursan, estado civil, discapacidad física, nivel económico, edad, sexo, trabajo del padre y la madre, notas de los exámenes, entre otros. A partir del árbol de decisión J48 de la herramienta Waikato Environment for Knowledge Analysis (WEKA) se obtiene datos que alerta al docente sobre los estudiantes en riesgo de desertar.

La presente investigación pretende diseñar un modelo para determinar factores de deserción estudiantil a través de técnicas de minería de datos para determinar su influencia predictiva en la deserción estudiantil universitaria en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

3 BENEFICIARIOS DEL PROYECTO

Los beneficiarios directos del proyecto son: los estudiantes, las autoridades y el personal docente de la Facultad de Ciencias de la Ingeniería y Aplicadas.

Los beneficiarios indirectos son: el personal administrativo y empleados de la Facultad de Ciencias de la Ingeniería y Aplicadas.

4 EL PROBLEMA DE INVESTIGACIÓN

La deserción estudiantil universitaria es un fenómeno que no respeta ninguna nación o continente, aunque en diversos países los índices varían de uno a otro. De acuerdo al reporte del Instituto Internacional para la Educación Superior en América Latina y el Caribe (IESALC), existen cuatro factores principales que influyen en la deserción: socioeconómicos, propio sistema universitario, orden académico y personales (Viteri Castro & Uquillas Narváez, 2011).

Con el entorno amplio de matrícula estudiantil, aparecieron las primeras investigaciones sobre el abandono universitario en Argentina, promovidos por los sociólogos Jorge Graciarena (1961), Gino Germani y Ruth Sautu (1965). Se dieron otras investigaciones desde otros campos disciplinarios como la economía que sea tanto local como internacional, en el cual se vincularon con diversas tradiciones teorías y disciplinarias (Santos Sharpe & Carli, 2016).

Por otra parte, la deserción en los estudios universitarios ha existido hace varias décadas atrás y sigue existiendo en la actualidad, esto se ha convertido en un problema grave en el ámbito de educación superior, debido a las altas tasas de deserción que se presenta en los estudiantes de primer año de carrera, quienes abandonan sus estudios en un 19% de acuerdo al Ministerio de Educación, Cultura y Deporte en España en el año 2015 (Tuero, Cervero, Esteban, & Bernardo, 2018).

El porcentaje de deserción universitaria a nivel mundial es del 50% de acuerdo al Banco Mundial de Estados Unidos en el año 2017, que indica que la mitad de los estudiantes no llegan a concluir sus estudios universitarios lo cual afecta a nivel económico, social y político (Breno, Peralta, Remache, & Mayorga, 2018). De acuerdo a los datos reportados por la Oficina de Planeación de la Universidad de Caldas en el periodo 2012-2013, el 30.0% de los estudiantes

abandonaron sus estudios por problemas económicos, el 25.9% por una mala elección de la profesión a seguir, el 0.8% por problemas institucionales, el 2.1% por motivos relacionados con el embarazo y el 0.7% por motivos de enfermedad (Gardner, Dussán, & Montoya, 2016).

El porcentaje de deserción universitaria en el Ecuador corresponde al 26%, según fuentes del informe de rendición de cuentas de la SENESCYT en el año 2015 y 2017. Las carreras con mayor porcentaje de deserción son: Ingeniería en Sistemas con el 65,45%, Ingeniería en Electrónica y Telecomunicaciones con el 55,36%, Ingeniería Industrial con el 57,14%, Ingeniería Eléctrica con el 14,38%, Ingeniería Mecatrónica con el 21,14%, Medicina Veterinaria y Zootecnia con el 26.09%, Ingeniería de Sistemas y Telemática con el 39,66%, Licenciatura en Arte Teatral con el 35,29% e Ingeniería Electrónica con el 35,09% (Sinchi, 2018).

Por otro lado, existe en la literatura varios factores que determinan la deserción universitaria. Estos factores pueden estar relacionados con el hecho de que los estudiantes llegan sin conocimientos, sin ninguna información, sin motivación y no cuentan con una planificación para su formación profesional (Pérez Álvarez & Aguilar López, 2017). Existen otras causas comunes de deserción, entre ellas se encuentran: las económicas, familiares o una mala elección de la profesión a seguir, sociales y académicas que depende de cada país, esto ocasiona un bajo rendimiento académico (Velasco Quintero, 2016).

El modelo de deserción propuesto por Vicent Tinto en 1975, estableció un análisis del compromiso del estudiante con la universidad a la cual ingresó, añadiendo los compromisos propios de sus metas académicas. El autor indica que ésto puede ser un determinante para persistir o abandonar la institución. Pues los compromisos planteados pueden verse afectados por factores del mismo estudiante y el entorno que lo rodea al ingresar en la universidad (Saldaña Villa & Barriga, 2010).

A pesar de que la literatura presenta varios estudios aplicados para estudiar la deserción en las universidades, los diferentes modelos para predecir corresponden a modelos teóricos, los cuales se enfocan en determinar las causas que pueden afectar al estudiante y por consecuencia llegar a tomar la decisión de permanecer o desertar de sus estudios académicos de la universidad.

Existen modelos que aplican técnicas de Machine Learning para predecir los estudiantes en riesgo de desertar, sin embargo, los modelos identificados fueron elaborados para un contexto de educación superior diferente al que se aplica en el Ecuador. Por lo tanto, se considera importante diseñar un modelo que permita conocer las características propias de los estudiantes

ecuatorianos, especialmente en carreras de formación de pregrado técnica en la provincia de Cotopaxi.

4.1 Definición del problema

¿Cuáles son los factores que influyen en la deserción de los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas?

5 OBJETIVOS:

5.1 General

- Diseñar un modelo para determinar factores de deserción estudiantil en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi, a través de la minería de datos.

5.2 Específicos

- Revisar sistemáticamente la literatura para conocer las concepciones teóricas y filosóficas de la minería de datos y su influencia en la predicción de la deserción universitaria.
- Construir un DataSet con información de los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas para conocer la percepción de los estudiantes respecto a la deserción estudiantil universitaria.
- Determinar la analítica de datos para la construcción del modelo de deserción a través del uso de minería de datos.
- Validar el modelo basado en minería de datos para determinar factores de deserción estudiantil, a través de métricas de precisión.

6 ACTIVIDADES Y SISTEMA DE TAREAS EN RELACIÓN A LOS OBJETIVOS PLANTEADOS

Tabla 1: Actividades de los objetivos

Objetivos	Actividad (tareas)	Resultado de la actividad	Descripción de la actividad (técnicas e instrumentos)
Revisar sistemáticamente la literatura para conocer las concepciones teóricas y filosóficas de la minería de datos y su influencia en la predicción de la deserción universitaria.	Búsqueda de información en las bases de datos científicas. Investigar las bases teóricas del proyecto. Análisis de los contenidos de los documentos identificados.	La identificación de estudios potenciales. La adquisición de información oportuna acerca del tema de investigación. Marco teórico.	Búsqueda de información en las bases de datos científicas, Google Scholar, Scielo, Redalyc, entre otros.
Construir un DataSet con información de los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas para conocer la percepción de los estudiantes respecto a la deserción estudiantil universitaria.	Realizar las preguntas de la encuesta. Validación y confiabilidad de la encuesta. Aplicación de las encuestas. Tabulación de los datos a través de SPSS.	Cuestionario. Análisis de fiabilidad de la encuesta. DataSet.	La técnica a utilizar será la encuesta, por medio de la cual se pretende adquirir y analizar la información del caso de estudio.
Determinar la analítica de datos para la construcción del modelo para determinar los factores de deserción a través del uso de minería de datos.	Selección. Pre-procesamiento (selección de variables). Extracción. Interpretación y evaluación.	DataSet con información confiable. Selección de las variables más significativas. Determinación de las tasas de precisión de la deserción.	Se aplicaron las técnicas de minería de datos que posee la herramienta Weka tales como: la red neuronal (Multilayer Perceptron), los árboles de decisión (J48 y Random Forest), métodos bayesianos (Naive Bayes) y Minimal Sequential Optimization (SMO).
Validar el modelo basado en minería de datos para determinar factores de deserción estudiantil, a través de métricas de precisión.	El resultado de las instancias correctamente clasificadas, el Kappa Statistic, el error absoluto medio, el F-Measure y la curva ROC.	Aplicación de métricas para evaluar los resultados obtenidos. Tasas de predicción.	Sensibilidad: para saber la proporción identificados de casos positivos de la predicción. Especificidad: para descubrir la proporción de casos negativos de la predicción identificados correctamente. Precisión: para conocer la exactitud del modelo de predicción.

Fuente: Grupo de trabajo

7 FUNDAMENTACIÓN CIENTÍFICO TÉCNICA

7.1 Antecedentes

Las primeras indagaciones sobre la deserción estudiantil, se extrajeron como base conceptual de la teoría del suicidio de Durkheim (1897), la perspectiva económica analiza los costos y beneficios de la educación. La deserción se la asemeja como un suicidio de la sociedad, los centros de educación superior son considerados sistemas con propios valores y estructura social por Spady (1970), en el que los niveles bajos de incorporación social permitan que crezca la deserción, siendo estas investigaciones para abordar dicho problema desde una perspectiva individual. Tal como los desarrolladores en la deserción se concentraron en la ampliación teórica del problema, pues existen estudios que se han interesado en las causas del fenómeno para obtener evidencias empíricas (Castaño & Vásquez, 2004).

7.2 Principales referentes teóricos

Según el autor (Miranda & Guzmán, 2017) en el análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos. La deserción estudiantil universitaria afecta a las universidades y estudiantes. Este autor define a la deserción como el abandono de los estudios antes de obtener el título. En el trabajo del autor se analizaron a 12 carreras comprendidas entre el año 2000 – 2013, con una muestra de 73.958 alumnos estudiados para lo cual se estableció valores estándares de 1 y 0 respetivamente.

En el mismo trabajo para determinar las variables que inciden a que un estudiante renuncie de sus estudios universitarios, se aplicó el análisis de minería de datos, estableciendo una metodología de tres algoritmos como son: redes bayesianas, redes neuronales, y árboles de decisiones, permitiendo que dichos algoritmos determinen si un estudiante desertará o no, por medio de la herramienta Waikato Environment for Knowledge Analysis (WEKA). Se utilizó el proceso Knowledge Discovery in Databases (KDD), que consiste en la fuente de información, preparación de datos, minería de datos, interpretación y evaluación de los conocimientos. Luego de aplicar los tres algoritmos, así como los softwares, Statistical Package for the Social Sciences (SPSS) y Waikato Environment for Knowledge Analysis (WEKA), se obtuvieron resultados en porcentajes que permitieron que los valores generados por los mismos fueran comparados, aunque en este caso no fueron totalmente comparables

Se estableció el análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. Del mismo modo el autor señala que la deserción está presente en el sistema educativo el mismo que está conectado con la forma de selección y

rendimiento académico. Es un tema que hace varios años es preocupante, por tal razón se realizan estudios para contribuir a determinar las causas que mayor efecto tienen en la deserción.

Por lado, se realizó un estudio en la carrera de Ingeniería en Informática de la Universidad de Gastón Dachary en Argentina desde el año 2000 – 2009, encontrando casos de 855 estudiantes, los factores estudiados fueron agrupados en Des (deserta) y NoDes (no deserta), mediante una técnica de minería de datos, se seleccionaron y depuraron datos para luego de ello aplicar algoritmos de clasificación como: árboles de decisiones, y redes bayesianas. Estos algoritmos fueron aplicados en la herramienta de Waikato Environment for Knowledge Analysis (WEKA), además que utilizaron otro algoritmo de Reglas OneR.

Tras la aplicación de los algoritmos se identificaron las variables que influyen en la deserción como es el caso de las asignaturas aprobadas, cantidad y resultado de asignaturas cursadas, procedencia y edad de ingreso del estudiante, siendo el caso de los estudiantes del primer año de carrera. Con todos los algoritmos aplicados, el resultado fue que los porcentajes eran similares y no se identificaron los mismos atributos (Eckert & Suénaga, 2015).

De igual manera, en el trabajo presentado por (Torres, Ramos, & Moraga, 2016) se realiza, el estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos, la tasa de deserción de los estudiantes de educación superior se utiliza para evaluar los procesos de enseñanza en las instituciones. La muestra para este caso fueron los estudiantes de primer año con 5.547 registros entre los años 2009 – 2015.

Para ello, se desarrolló el proceso de minería de datos con la metodología Cross Industry Standard Process for Data Mining (CRISP-DM), que tiene los siguientes pasos: comprensión del negocio, datos, muestra, preparación de datos, eliminar valores nulos y perdidos, transformar atributos y seleccionar variables. Se aplicó los algoritmos de árboles de decisiones, métodos bayesianos y redes neuronales.

Realizaron una evaluación con cuatro experimentos: el primero se utilizó todas variables, el segundo se descartó variables académicas, el tercero se omitió las variables de aprobación y promedio final y el cuarto solo se utilizaron las variables académicas de ingreso, las mismas que fueron analizadas. Los resultados fueron que en los experimentos las variables académicas no afectan a la deserción, las variables del desempeño académico tienden a afectar en la

deserción, con los resultados obtenidos de las técnicas de Random Forest y el Naive Bayes podrán intervenir en las variables que causen deserción en los estudiantes.

7.3 Bases teóricas

7.3.1 Deserción Universitaria

Es la renuncia decisiva del aula de clases por diversos motivos, la deserción conlleva un proceso lento que se encuentra en el interior del individuo y va creciendo hasta llegar a una decisión para bien o mal del mismo y el entorno en el que se encuentra (Yepes Delgado, Salazar Béltran, & Ceferino Martínez, 2017).

7.3.2 Modelo para Predecir

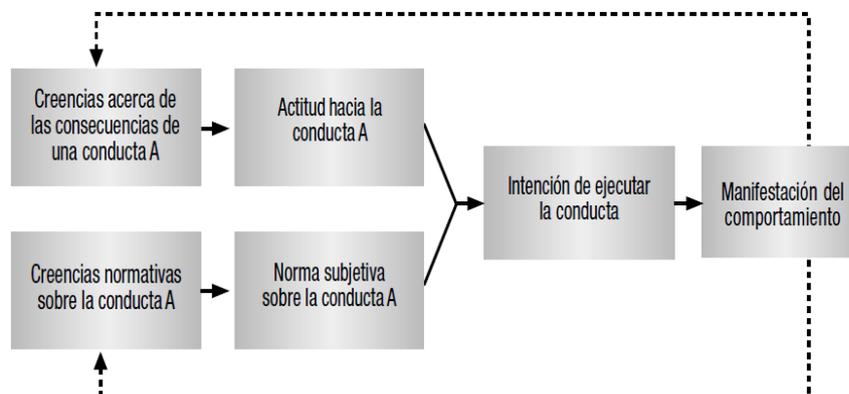
Existen modelos predictivos que han sido utilizados como base en la deserción estudiantil entre los cuales se encuentran los siguientes:

7.3.2.1 Modelo de Martin Fishbein e Icek Ajzen

Planteado en 1975 enfocándose en el campo de la psicología, en la que la conducta interviene por las creencias y actitudes del estudiante, con ello la decisión de desertar o continuar con los estudios se encuentra influenciada por varios factores considerados los más importantes: las conductas, posición sobre desertar o continuar y acciones que conllevan normas subjetivas (Chisco Ruiz, 2015).

La deserción es concebida de acuerdo a los resultados de la debilidad en los fines iniciales y perseverancia reforzando la misma. El pensamiento explica la deserción voluntaria y retención (Himmel, 2018). El modelo identificado se presenta en la Figura 1.

Figura 1: Modelo de Fishbein y Ajzen, 1975

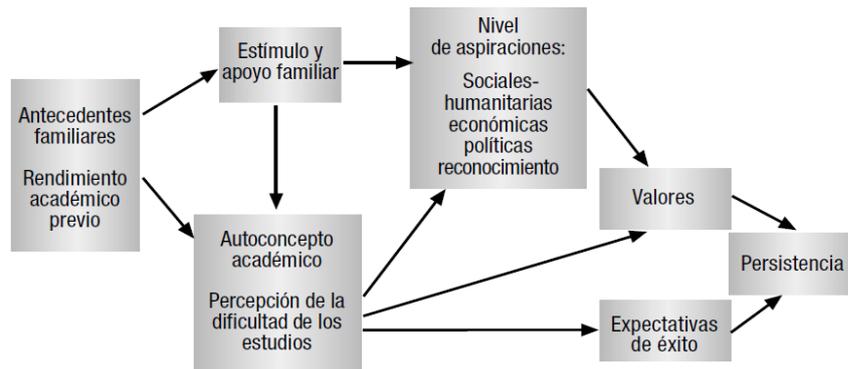


Fuente: Tomado de Fishbein y Ajzen (1975)

7.3.2.2 Modelo de Corinna Ethington

Este modelo propuesto en 1990, tiene una completa distribución de las conductas de logro. Se incluye la persistencia, elección y desempeño. El nivel de deseos tenía un resultado inmediato en los valores y posibilidades de éxito que influyen en la permanencia en la universidad (Chisco Ruiz, 2015). El modelo identificado se presenta en la Figura 2.

Figura 2: Modelo de Ethington, 1990



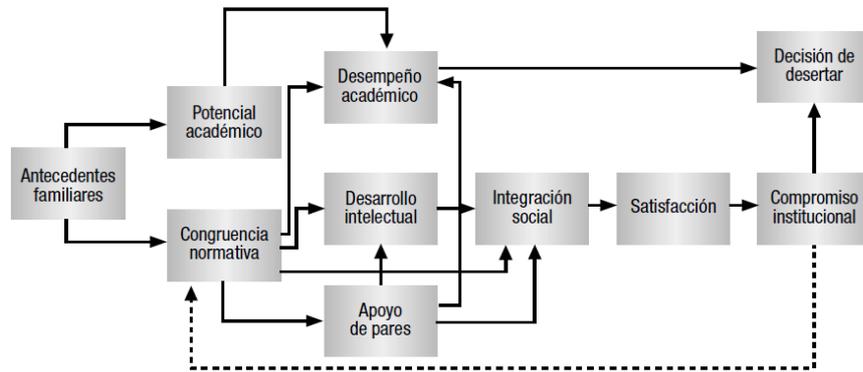
Fuente: Tomado de Ethington (1990)

7.3.2.3 Modelo de William Spady

Spady 1970 analiza la deserción, en donde se utiliza los principios de las causas de deserción, estableciendo que la decisión de abandonar los estudios depende de los factores individuales que tiene el individuo sino también de factores sociales pues existe una ruptura entre los mismos lo que impide que se relacione al individuo con la sociedad (Vásquez Verdugo, 2016).

Para el autor el entorno de la familia influye en la persona sobre su potencial en sus estudios, el rendimiento académico e intelectual, apoyo de pares y la integración social actúa sobre una lógica normativa. La integración social tiene un apoyo de pares que causan impactos de satisfacción en el estudiante e intervienen en la inserción en la universidad y favorecen a ratificar compromisos que el estudiante se propuso (Chisco Ruiz, 2015). El modelo identificado se presenta en la Figura 3.

Figura 3: El modelo de Spady, 1970

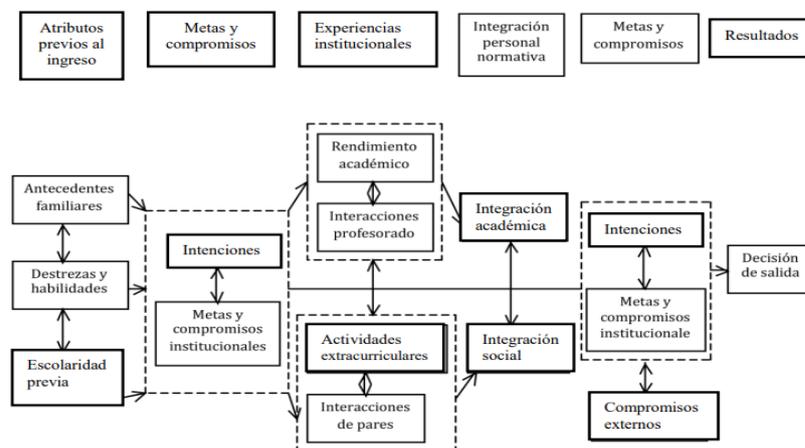


Fuente: Tomado de Spady (1970)

7.3.2.4 Modelo de Vicent Tinto

En el modelo de Tinto en 1975, citado por (Vásquez Verdugo, 2016) se establece que los estudiantes estarán en un programa en el que se registraron y cuando los beneficios sean superiores al esfuerzo, dedicación y otros costos de actividades en la que sus beneficios sean mayores, existe la probabilidad de que un estudiante tome la decisión de desertar. El modelo identificado se presenta en la Figura 4.

Figura 4: Modelo de Tinto, 1987, Adaptado de Tinto (1975)



Fuente: Tomado de Tinto (1975)

7.3.2.5 Modelo de John Bean

Bean complementa los modelos de Tinto y Spady en 1980 y 1985 por medio de dos trabajos, en el primer trabajo al existir evidencia empírica se puede testear mediante variables y en el segundo trabajo añade en el análisis a estudiantes no habituales como respuesta en aquella época a un cambio para poder acceder a la educación superior pues la misma era restringida a un grupo superior (Vásquez Verdugo, 2016). El modelo identificado se presenta en la Figura 5.

En la Tabla 2, se definen variables antecedentes que se encuentran relacionadas con el estudiante entre ellas están el estado socioeconómico, desempeño académico anterior y domicilio actual, dichos factores causaran impacto en las organizaciones o instituciones en las que se tome la decisión de desertar.

Tabla 2: Lista de variables planteadas por Bean 1980

Variable	Definición
Variables de Antecedentes	
Desempeño Previo	Grado en que el estudiante ha demostrado sus logros académicos previos.
Estatus Socioeconómico	Grado en que los padres del estudiante han logrado como estatus a través de la ocupación familiar.
Residente en el Estado	Si el estudiante es un residente del estado en donde la institución educacional está.
Distancia a Casa	Distancia de su residencia actual a la casa de sus padres.
Tamaño de la ciudad	Tamaño de la comunidad donde el estudiante pasó la mayor parte de su tiempo en su crecimiento.
Determinantes Organizacionales (Basado en (Price, 1977))	
Rutina	El rol de ser un estudiante es visto como una rutina.
Desarrollo	Desarrollo académico.
Valor Práctico	Percepción del estudiante.
Calidad Institucional	Nivel de participación del estudiante
Integración	En relaciones primarias o cuasiprimarias (tiene amigos cercanos).
Promedio de Notas Universitario	Certificaciones obtenidas por el estudiante.
Compromiso de Metas	Nivel de metas académicas cumplidas.
Comunicación	Comunicación del estudiante.
(Requerimientos/Reglas) Justicia Distributiva	El estudiante cree que es tratado justamente por la institución. Por ejemplo: recibe premios y castigos proporcionalmente a su esfuerzo realizado en su rol como estudiante.
Centralización	El estudiante cree que participa en los procesos de tomas de decisión. Por ejemplo: centros de estudiantes, consejeros, etcétera.
Advisor	El estudiante cree que su advisor es útil.
Relación con funcionarios	Nivel de contactos informales con los miembros de la facultad.
Trabajo en el Campus	Necesidad de tener un trabajo en el campus universitario para permanecer en la escuela.
Mejor (área)	El área de uno de los campos de estudio.
Mejor (certeza)	El estudiante es poco indeciso en que se está especializando.
Alojamiento	Cuando una persona vive On Campus
Organización del Campus	El número de miembros en la organización del campus.

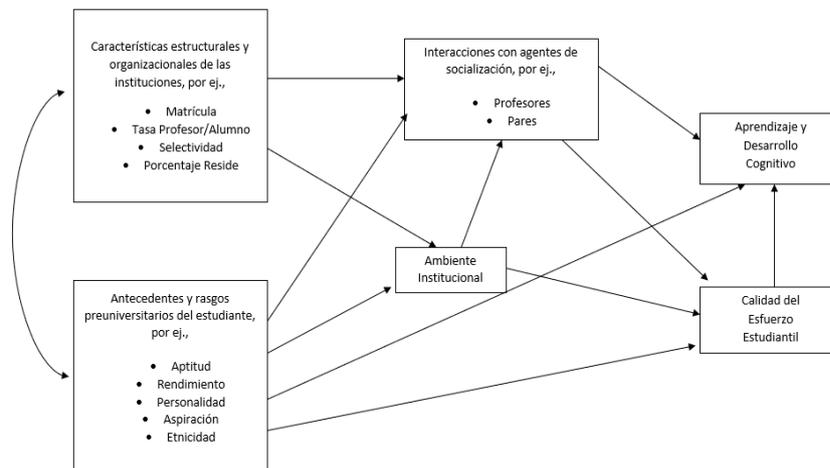
Fuente: Adaptado de Vásquez (2016)

Figura 5: Modelo de Bean, 1985

Fuente: Tomado de Modelo de Bean (1985)

7.3.2.6 Modelo de Ernest Pascarella y Patrick Terenzini

Según (Chisco Ruiz, 2015) el modelo de Pascarella y Terenzini (1985) “propone un modelo causal en relación con las características institucionales y ambientales. Este es concurrente con estudios de impacto en diversas universidades”. El modelo identificado se presenta en la Figura 6.

Figura 6: Modelo de Pascarella y Terenzini

Fuente: Adaptado de Chisco (2015)

7.3.3 Tipos de deserción

La deserción se da en dos tipos con respecto al tiempo y espacio.

➤ Tiempo

- Deserción precoz: Persona que fue aceptada en la institución superior y no se matricula.

- b) Deserción temprana: Persona que una vez iniciado con sus estudios académicos decide abandonar los estudios en un lapso de los dos primeros semestres (Fontalvo Cerpa, Castillo González, & Castillo, 2015).
 - c) Deserción tardía: Persona que renuncia a sus estudios en los últimos niveles de estudio (Corzo Salazar, 2017).
- Espacio
 - a) Deserción institucional: El estudiante deja la universidad (Páramo & Correa, 1999).
 - b) Deserción interna o del programa académico: El estudiante que opta por cambiarse de carrera en la misma universidad (Gómez Torres, Acevedo Villalobos, & Salamanca Velandia, 2015).

7.3.4 Factores de deserción

- Factores externos
 - a) Factores institucionales: Describe las oportunidades y posibilidades que la universidad ofrece para comenzar o continuar con sus estudios.
 - b) Factores organizacionales: Servicios de la institución que esta ofrece a los estudiantes que ingresan, enfocándose en la calidad de docencia y experiencias en las aulas.
- Factores bilaterales
 - a) Factores personales: Depende del estudiante y están relacionados por motivos psicológicos como emociones, motivos sociológicos debido a influencias familiares. La insatisfacción de expectativas que desean cumplirse, así como la falta de motivación para lograr sus objetivos.
 - b) Factores socioculturales: La forma en como interactúa en el medio que se encuentra. Inclusive con la integración en un grupo de personas de su misma edad y el ambiente universitario que pueda generarse cuando un estudiante comienza sus estudios.
- Factores internos
 - a) Factores académicos: El bajo rendimiento que un estudiante puede tener, así como las dificultades de aprendizaje. Las formas de estudio seguí para obtener resultados que sean satisfactorios (Zavala-Guirado, Álvarez, Vázquez, Gonzáles, & Bazan Ramírez, 2018).

7.3.5 Minería de datos

Se refiere a la obtención de la información que se encuentra oculta, en grandes bases de datos almacenados en distintos formatos, es decir un proceso posterior para descubrir conocimiento útil de la información disponible, a fin de contribuir a la toma de decisiones mediante la generación de información y el análisis de la misma, permitiendo a los negocios tomar decisiones proactivas (Echeverry Marulanda, Trujillo López, & Mejía Salazar, 2017).

7.3.6 Las técnicas de minería de datos (Data Mining)

Cuando las herramientas de Data Mining son implementadas en el software de procesamiento de alta performance, permite analizar los datos almacenados en una base de datos en cuestión de minutos, haciendo que el usuario pueda analizar dichos datos y a su vez producen mejores predicciones (Presser Carne, 2009).

7.3.6.1 Clasificación de las técnicas de minería de datos:

Supervisadas o predictivas

Se utilizan para predecir valores o conducta futura de algún tipo de entidad a través de las variables dependientes, cuya base se encuentra en las variables independientes previamente relacionadas (Molina López & García Herrero, 2006).

➤ Predicción

Proceso que busca determinar los valores de una o varias variables, a partir de un conjunto de datos. La predicción de valores continuos puede planificarse por las técnicas estadísticas de regresión (Chamba Jiménez, 2015).

➤ Regresión

La regresión logística, brinda la posibilidad de evaluar la influencia de todas las variables independientes en relación a la variable dependiente, así como controlar el efecto del resto, se puede tomar únicamente dos valores ya sea (0) si el hecho no ocurre o (1) si el hecho ocurre, a este proceso se le llama binomial ya que solo presenta dos resultados posibles, siendo la probabilidad de cada uno de ellos constante en una serie de repeticiones (Holgado Apaza, 2018).

➤ Clasificación

La clasificación supervisada busca un modelo válido que permita predecir casos futuros por medio del aprendizaje de casos conocidos, el proceso consiste en distribuir un conjunto de datos en grupos, donde cada miembro de un grupo esté cerca de otros y grupos diferentes estén lejos

de otros, donde la distancia se mide con relación a las variables especificadas, que se quieren predecir (Chamba Jiménez, 2015).

Árboles de decisión

Es un modelo predictivo con facilidad para interpretar el modelo generado debido a su capacidad explicativa, consta de nodos que hace referencia a regiones clasificadas (etiquetadas) de una clase, y nodos internos (splits) que representan alternativas para seleccionar la subregión al que se dirige el elemento que llega al nodo, es decir cuando ingresa un nuevo dato al árbol de decisión inicia su recorrido por el nodo raíz, el mismo que posee una condición en el que especifica por qué rama del árbol debe descender, repitiendo este proceso hasta llegar a la hoja donde se toma una decisión final (Gironés Roid, Casas Roma, Miguinllón, & Caihuelas, 2017).

- Random forest

Es un algoritmo que contiene una combinación de técnicas de Classification And Regression Tree (CART) y Bootstrap Aggregation (Baggin), lo que permite obtener árboles predictores en la que cada árbol se crea a partir de valores que son proporcionados por vectores aleatorios independientemente comprobado y con una distribución igualitaria. Con este algoritmo se obtiene una mejor precisión de clasificación al construir cada uno de los clasificadores individualmente de forma aleatoria (Brian, Flores, & Ponce, 2018).

El Baggin es una técnica diseñada para la creación de datos de entrenamiento que remuestrea aleatoriamente el conjunto de datos originales con reemplazamiento, es decir, sin eliminar el conjunto de datos seleccionados de forma previa a la elección del siguiente subconjunto. Por tanto, algunos datos pueden ser usados más de una vez en el entrenamiento de los clasificadores individuales (Rodríguez & Chica, 2012).

- Naive bayes

Se basa con métodos estadísticos de probabilidad que manifiestan la probabilidad condicionada de cada clase, dada las propiedades de un ejemplo (atributos), seguidamente los parámetros suelen ser combinados para designar las clases que maximizan sus probabilidades a nuevos ejemplos (Gironés Roid et al., 2017).

- Redes de neuronas

Las redes neuronales artificiales son algoritmos adaptivos y elaborados debido a su capacidad de aprendizaje, es decir una agrupación de algoritmos de comunicación de la neurona biológica

permitiendo resolver problemas relacionados con la clasificación y predicción. Las redes neuronales vinculan entradas con salidas, generando aproximaciones universales debido a su capacidad de aprender y aproximar con precisión la función $f(x) = y$. Donde X = datos de entrada y $F(x)$ = resultados del algoritmo (Gironés Roid et al., 2017).

- Multilayer perceptron

Se define como una arquitectura de redes neuronales teniendo la posibilidad de generalizar para generar resultados de forma rápida, posee características de V4 (volumen, velocidad, variedad y variabilidad) que corresponden a los datos. Se emplean estimadores actualizados a diferencia de estimadores fijos, que hacen que el modelo pierda precisión luego de un cierto tiempo y se produzca errores significativos. Este perceptron puede reconocer parámetros que necesiten de una actualización cada cierto tiempo permitiendo que se modifique solo los estimadores necesarios para que sea un modelo más preciso que al usar estimadores dinámicos y fijos (Climent Hernández, Hoyos Reyes, & Martínez Preece, 2017).

No supervisadas o descriptivas

La técnica descriptiva comprende el aprendizaje sin supervisión, permitiendo generar patrones que explican la estructura interna, las relaciones y la conexión recíproca con los datos extraídos (Molina López & García Herrero, 2006).

➤ Clustering

Se refiere a la agrupación de datos en clases o clusters que permita a los objetos de un cluster mantener una similitud alta entre sí, y una similitud baja (sean diferentes) con objetos de otros clusters, es una técnica de aprendizaje automático no supervisado que conlleva a la búsqueda de una estructura en una base de datos no etiquetados (Holgado Apaza, 2018).

➤ K – means

El algoritmo K – means, es uno de los métodos de clustering, que representa el agrupamiento de datos por vecindad, parte de un número determinado de prototipos y de un conjunto de ejemplos a agrupar sin etiquetar, con el objetivo de identificar a los prototipos de manera que los datos pertenecientes al mismo prototipo posean características similares (Alania Ricaldi Pit Frank, 2018).

➤ Asociación

Realiza un análisis exploratorio a fin de encontrar relaciones dentro del conjunto de datos independientes (Chamba Jiménez, 2015).

➤ Minimal Sequential Optimization (SMO)

Implementa el algoritmo de optimización mínima secuencial para entrenar una máquina de soporte vectorial (SVM), en el que transforma los datos originales de forma no lineal en un nuevo espacio de mayor dimensión. En este nuevo espacio se construye un modelo lineal que pueda representar un límite de decisión no lineal en el espacio original (Martínez & Sanjurjo, 2014). Es decir, con la implementación del algoritmo Minimal Sequential Optimization (SMO) se reemplaza globalmente todos los valores faltantes y transforma los atributos nominales en valores binarios, también normaliza todos los atributos por defecto. En ese caso, los coeficientes en la salida se basan en los datos normalizados, no en los datos originales (Weka, 2019).

7.3.7 Campos de aplicación de la minería de datos en el Ecuador

7.3.7.1 Meteorología

Un proyecto aplicando la minería de datos permitió obtener datos de patrones de comportamiento meteorológico, mediante los datos que posee el Observatorio Astronómico de Quito (OAQ), recopilados durante 10 años. Como resultado de un patrón de comportamiento se obtuvo que la humedad relativa disminuye cuando la temperatura aumenta (Camana, 2016).

7.3.7.2 Elecciones presidenciales

Al utilizar herramientas de minería de datos, para la selección de juntas receptoras del voto (JRV) más representativas, se aplicó la técnica de agrupamiento Clustering, el algoritmo utilizado fue el K-MEDIAS. Para el presente estudio se eligieron 20 clusters, donde se determinó que en el cluster 13 se ubican las Juntas Receptoras del Voto con resultados más cercanos a la distancia Euclídea (Camana, 2016).

Para la distribución del voto se utilizó el árbol de decisión, ya que al ser intuitivos y fáciles de comprender permitieron obtener información de las parroquias, cantones y provincias que tienen alta representación de Juntas Receptoras del Voto. Los datos obtenidos sirven de apoyo para conocer los posibles resultados y a la vez que sirven de base para las próximas elecciones (Camana, 2016).

7.3.7.3 Educación

En la Universidad Nacional de Loja, se aplicó la minería de datos para determinar las interacciones de los estudiantes del curso virtual del idioma inglés, de la modalidad de Estudios a Distancia (MED), correspondiente al periodo académico 2013-2014. Por medio del algoritmo árbol de decisión se determinó que los factores que más inciden son: las tareas, exámenes, recursos, estado civil y la situación laboral del estudiante (Camana, 2016).

7.3.8 Ventajas y desventajas de la minería de datos

Según (Equipo Editorial, 2018) existen ventajas y desventajas de la minería de datos que serán expuestas a continuación:

➤ Ventajas

- Previene un entorno desfavorable futuro mostrando datos verídicos.
- Con la información obtenida se puede asistir en la toma de decisiones de estrategias.
- Los usuarios pueden comprender de mejor manera la información y concomimientos.
- Permite crear formas de enseñanza para mantener clientes, disminuyendo el riesgo de perder usuarios.

➤ Desventajas

- Al contar con una gran intensidad de trabajo se puede necesitar de una inversión para los equipos y capacitación personal.
- Existe una inseguridad en los usuarios estudiados por la información que proporcionada para el estudio.
- Para procesar la información puede tomar algo de tiempo de acuerdo a la cantidad de base de datos.
- Al no contar con deficiente sistema de seguridad la información del usuario se pondría en riesgo.
- Cuando la información es inexacta, el proceso no es cumplido y afectaría en la toma de decisiones de los resultados.

7.3.9 Las herramientas de minería de datos

Las herramientas de la minería de datos se refieren al software eficiente, a través del cual se puede buscar y analizar la información de forma rápida y oportuna, optimizando el tiempo de respuesta.

➤ Waikato Environment for Knowledge Analysis (WEKA)

Es un software de minería de datos libre, con una multiplataforma desarrollada por Java en la Universidad de Waikato, contiene una variedad de algoritmos, sus archivos tienen un propio formato arff (Attribute Relation File Format) cuenta con una interfaz fácil de manejar, pero al realizar un proceso en casada no es fácil debugearlo (Lanzarini & Villa Monte, 2017).

Waikato Environment for Knowledge Analysis (WEKA) tiene algoritmos de clasificación, los métodos estadísticos usados para validar la clasificación y métodos en las que se puede disminuir, eliminar transformar datos, tareas de clasificación, regresión clustering (Villanueva Morales, Lugo Rodríguez, Ramírez Buenrostro, & Ramírez Pérez, 2015).

➤ Statistical Package for the Social Sciences (SPSS)

La herramienta Statistical Package for the Social Sciences (SPSS) fue creada por International Business Machines (IBM), es un conjunto de herramientas utilizadas para un análisis estadístico, mediante el procedimiento de los datos. Utiliza un soporte operativo de Windows con menús desplegables y cuadro de diálogos que facilitan el trabajo del mismo (Guardiola Jiménez, 2019).

Según (Universidad Privada TELESUP, 2019) muestra las características y ventajas del Statistical Package for the Social Sciences (SPSS):

Características del Statistical Package for the Social Sciences (SPSS)

Este tipo de software es el más utilizado en todo el mundo permitiendo una investigación, experimentación y toma de decisiones.

- Se puede realizar operaciones aritméticas, algebraicas y trigonométricas en las hojas de cálculo.
- Procesa los datos de forma dinámica del gestor de base de datos.
- Permite realizar informes luego de una investigación ejecutada.
- Efectúa búsquedas inteligentes de un gestor de minería de datos para extraer información de árboles de decisión, segmentaciones del mercado o redes neuronales.

Ventajas del Statistical Package for the Social Sciences (SPSS)

- Permite ahorrar tiempo y esfuerzo.
- Cálculos más exactos.

- Trabaja con grandes cantidades de datos.
- Compatible con programas del manejo de datos.

➤ Eviews

Es un software para el análisis estadístico y econométrico de los datos, se utiliza para el estudio de series temporales, series atemporales y datos de panel. Eviews suministra un sofisticado análisis de los datos, modelos de regresión y predicción. El software incluye la tecnología de hoja de cálculo básica para presentar y analizar datos con los principales análisis estadísticos y econométricos. Para ello incluye una interfaz gráfica de usuario, pero además incluye una ventana de comandos que permite implementar cualquier tipo de análisis de datos con la codificación adecuada (Barreiro, Vicente, & Pintos Clapés, 2005).

7.3.10 Metodología de minería de datos

➤ Knowledge Discovery in Databases (KDD)

Este proceso iterativo e interactivo en la que el problema con técnicas de análisis de datos tradicionales y tecnológicas avanzadas para aprender (Gervilla García et al., 2016). El análisis de este proceso conlleva a una solución de un gran grupo de datos permitiendo obtener conocimientos mediante la información útil y a la vez permite asociar patrones válidos de los datos, es un proceso que tiene seis etapas.

- Pre-etapa de Identificación de objetivos: En esta etapa de Knowledge Discovery in Databases (KDD) se analizan los objetos que tendrá la misma, desde la perspectiva del cliente para obtener un conocimiento anticipado del negocio y los objetivos sean claros.
- Etapa de Selección: Se extraen los datos importantes luego proceder al análisis.
- Etapa de Preprocesamiento: Es la preparación y limpieza de datos obtenidos de un DataSet, eliminando inconsistencias o redundancia de información mediante técnicas y operaciones.
- Etapa de Transformación: Los datos que se ingresaran a un algoritmo de minería de datos debe ser apropiados.
- Etapa de Extracción: Se origina el conocimiento en el que se hallan las relaciones que existen entre los datos y patrones no conocidos, válidos, nuevos y útiles.
- Etapa de Interpretación y Evaluación: Se usan técnicas de evaluación para descifrar y comprobar el resultado de los patrones, los mismos que deben tener tres cualidades

precisos, comprensibles e interesantes que sean de utilidad y novedoso (Fernández et al., 2018).

➤ Cross Industry Standard Process for Data Mining (CRISP-DM)

Sus creadores y la experiencia que tienen establecieron un enfoque práctico, una asociación de empresas que incluyen software estadístico, Statistical Package for the Social Sciences (SPSS) de International Business Machines (IBM), es usada en minería de datos para el desarrollo de proyectos. Comprende seis fases que son (Oviedo C., Oviedo C., & Vélez S., 2015).

- Entendimiento de negocio: Establece los objetivos de la minería de datos y desarrollo del plan del proyecto.
- Entendimiento de datos: Recolectan y verifican los datos.
- Preparación de datos: Identificar, seleccionar, limpiar, construir e integrar los datos.
- Modelado: Seleccionar y evaluar el modelo.
- Evaluación y despliegue: Evaluar los resultados, revisar el proceso, planeación de implementación, plan de seguimiento, elaboración del reporte (Tamayo Benalcázar, 2017).

➤ SEMMA (Sample, Explore, Modify, Model, Assess)

La metodología descubre patrones de negocio desconocidos por medio del proceso de selección, exploración y modelado de los volúmenes de datos. Está orientada a los aspectos técnicos sin tomar en cuenta los análisis y comprensión del problema, utiliza un software de minería de datos de la compañía Statistical Analysis Systems (SAS) para trabajar y a la vez esta metodología contiene un conjunto de herramientas. SEMMA (Sample, Explore, Modify, Model, Assess) está conformado de cinco fases las mismas que se relacionan por su acrónimo en inglés:

- Muestreo: Es una parte que se extrae de un conjunto de datos que se pueda operar rápidamente y a la vez obtener información importante.
- Exploración: Busca tendencias e incoherencias inesperadas de los datos para comprender la información que se manipula.
- Modificación: Se cambian los datos mediante la creación, selección y evolución de las variables orientadas en el proceso de selección correcta del modelo.
- Modelado: El software realiza una combinación de datos para predecir un resultado, realizándolo de manera automática.

- Valoración: Los datos son evaluados por medio de su utilidad y fiabilidad luego de haber llegado a las conclusiones del proceso (Tamayo Benalcázar, 2017).

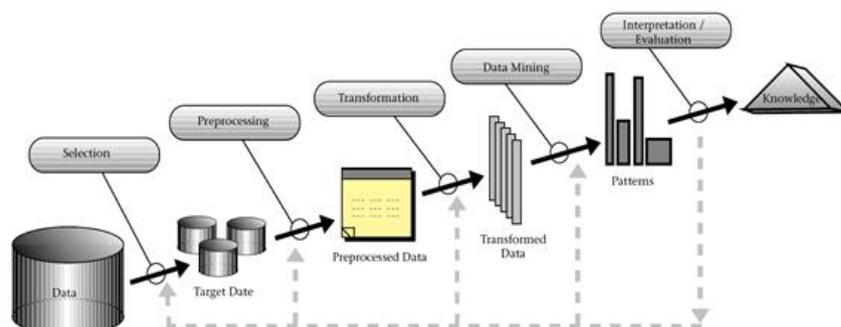
Tabla 3: Cuadro comparativo de las metodologías de minería de datos

CUADRO COMPARATIVO			
Metodología	Knowledge Discovery in Databases (KDD)	Cross Industry Standard Process for Data Mining (CRISP-DM)	SEMMA (Sample, Explore, Modify, Model, Assess)
Fases	Consta de 5 fases: selección, preprocesamiento, transformación, extracción y evaluación e implantación.	Consta de 5 fases: comprensión del negocio, entendimiento de datos, preparación de datos, modelado, evaluación y despliegue.	Consta de 5 etapas: muestreo, exploración, modificación, modelado y valoración.
Pasos	Los que se destaca la limpieza y preprocesamiento de datos, análisis exploratorio de selección del algoritmo de minería de datos, entre otros.	Ciclo de vida de acuerdo al modelo de referencia, de acuerdo a tareas.	Se extrae la población muestral, se realiza una exploración de la información sobre los datos obtenidos.
Proceso	Trabaja con un proceso iterativo-incremental.	Es un enfoque práctico.	Descubre patrones de negocio desconocidos por medio del proceso de selección, exploración y modelado de los volúmenes de datos.

Fuente: Tomado de Gervilla (2016), Oviedo & Vélez (2015) y Tamayo (2017).

Luego de analizar cada una de las metodologías de minería de datos Knowledge Discovery in Databases (KDD), Cross Industry Standard Process for Data Mining (CRISP-DM) y SEMMA (Sample, Explore, Modify, Model, Assess), la metodología que se utilizará es la Knowledge Discovery in Databases (KDD) porque está conformada por un proceso iterativo - incremental, el mismo que consta de una pre-etapa y cinco etapas, en el caso de no obtener resultados que sean satisfactorios al concluir las etapas, se puede volver a una etapa previa, de esta manera se llega a tener valores de datos que el investigador podrá usar para un posterior análisis.

Figura 7: Pasos que componen la metodología KDD



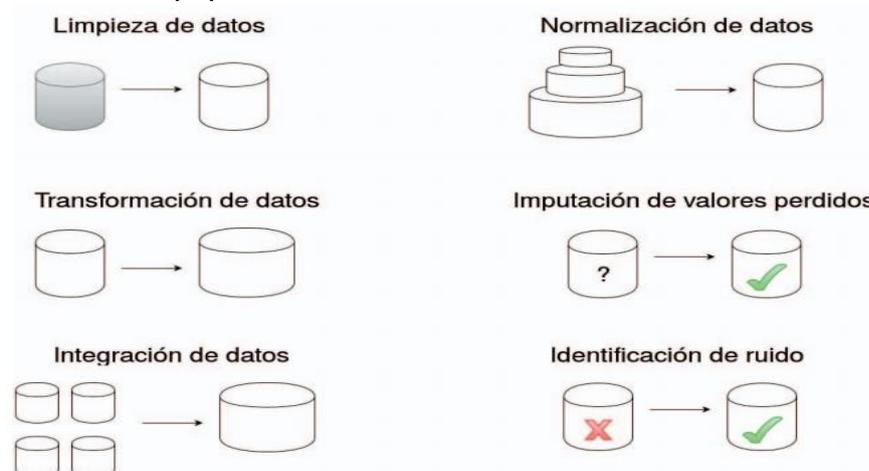
Fuente: Tomado de Fayyad (1996)

7.3.11 Técnicas de Preprocesado

El pre-procesamiento de datos es una de las etapas del proceso de descubrimiento de información o Knowledge Discovery in Databases (KDD), la cual se encarga de la limpieza de datos, de la integración, transformación y reducción para continuar con la siguiente etapa que es la extracción. Puesto que muchas de las veces los datos obtenidos contienen registros nulos o erróneos que conlleva un proceso de minería de datos con pobres resultados se aplica técnicas de preprocesamiento (García, Ramírez, Luengo, & Herrera, 2016).

El preprocesamiento incluye un rango amplio de técnicas que se agrupan en dos áreas: preparación de datos y reducción de datos. La preparación de datos se compone por una serie de técnicas cuyo objetivo es inicializar correctamente los datos que servirán de entrada para los algoritmos de minería de datos, por ello se considera importante incluso obligatorio el uso de estas técnicas ya que sin ellas los algoritmos de extracción del conocimiento no podrían ejecutarse y presentarían resultados erróneos, para ello se incluye la transformación de datos y normalización, integración, limpieza de ruido e imputación de valores perdidos (García et al., 2016).

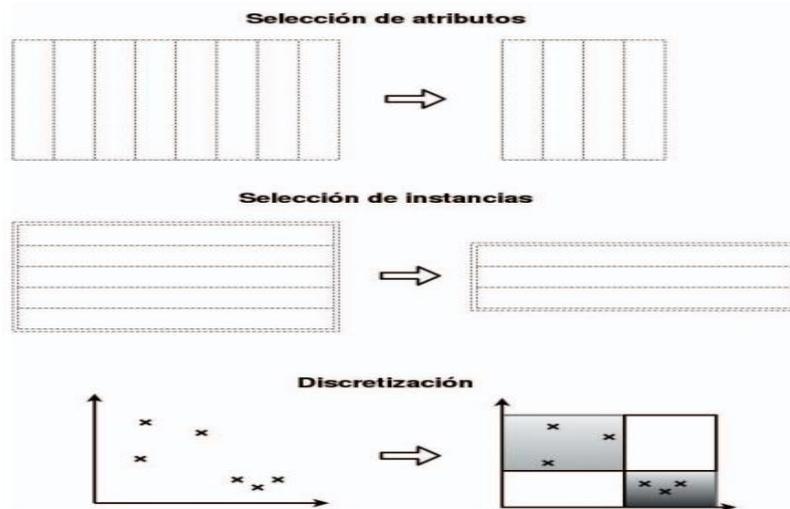
Figura 8: Familias técnicas en preparación de datos



Fuente: Tomado de García, Ramírez, Luengo & Herrera (2016)

Las técnicas de reducción de datos, se enfocan en obtener una representación reducida de los datos originales, manteniendo la integridad y la información existente en los datos. Las técnicas de reducción más relevantes son: la selección de atributos, la selección de instancias o la discretización (García et al., 2016).

Figura 9: Familias técnicas en reducción de datos



Fuente: Tomado de García, Ramírez, Luengo & Herrera (2016)

7.3.11.1 Normalize

La herramienta Waikato Environment for Knowledge Analysis (WEKA) posee diversos filtros sobre datos, permitiendo realizar transformaciones sobre ellos de todo tipo, entre estos tenemos Normalize: Normaliza todos los datos de manera que el rango de los datos pase a ser [0,1] (García Morate, 2008).

7.3.11.2 NumericToNominal

Un filtro para convertir los atributos numéricos en nominales. A diferencia de la discretización, solo toma todos los valores numéricos y los agrega a la lista de valores nominales de ese atributo. Útil después de las importaciones de Comma-Separated Values (CSV), para forzar que ciertos atributos se vuelvan nominales, por ejemplo, el atributo de clase, que contiene valores de 1 a 5 (García Morate, 2008).

7.3.12 Métricas para la evaluación de clasificadores

Las métricas empleadas para la evaluación de clasificadores son las técnicas de evaluación de hipótesis, basadas en precisión como, por ejemplo, el porcentaje de Error de Muestra (o inversamente el porcentaje de Acierto (Accuracy)), o el Alcance, la Precisión o la Especificidad (Recall, Precision, Specificity) o técnicas más elaboradas como el Área Bajo la Curva (ROC), si la curva ROC es = 1, el modelo es ideal para el conjunto de datos, si es mayor que 0.5 es óptimo y si es menor a 0.5 el modelo es despreciable (Haro, Pazmiño Maji, Conde, & Peñalvo, 2018).

Las métricas pueden ser calculadas en base al conjunto de datos que se utiliza para el entrenamiento y prueba, además se puede emplear la validación cruzada (o no), para evitar sesgo en las particiones de los datos (Dasso & Funes, 2012).

La precisión global (exactitud), muestra el número total de predicciones correctas del total, la precisión positiva (sensibilidad), proporción identificados de casos positivos, y la predicción negativa (especificad), proporción de casos negativos identificados correctamente, identificados previamente en la matriz de confusión (Dasso & Funes, 2012).

Según el autor (Bigdate, 2019) las métricas de evaluación son las siguientes:

$$P = \frac{TP}{TP+FP} \quad (1)$$

Precisión (P), alta significa pocos falsos positivos, para determinar la tasa de precisión de las técnicas de minería de datos se considera las ecuaciones propuestas por Bigdate en el año 2019, las cuales se representan en las ecuaciones (2), (3), (4).

Recall (sensibilidad, exhaustividad):

$$R = \frac{TP}{TP+FN} \quad (2)$$

Recall (R), alta significa pocos falsos negativos.

Tasa FP:

$$FPR = \frac{FP}{TN+FP} \quad (3)$$

Medida:

$$F1 = 2x \frac{PxR}{P+R}, \text{ es la media armónica de P y R.} \quad (4)$$

en donde, según Bigdate en el año 2019 corresponde a:

True Positives (TP): cuando la clase real del punto de datos era 1 (Verdadero) y la predicha es también 1 (Verdadero).

Verdaderos Negativos (TN): cuando la clase real del punto de datos fue 0 (Falso) y el pronosticado también es 0 (Falso).

False Positives (FP): cuando la clase real del punto de datos era 0 (False) y el pronosticado es 1 (True).

False Negatives (FN): Cuando la clase real del punto de datos era 1 (Verdadero) y el valor predicho es 0 (Falso).

Según el autor (Ruiz Reina, 2018) menciona que un clasificador que todo lo predijera positivo, obtendría $R=1$, pero habría muchos falsos negativos y P sería muy baja.

Un clasificador que sólo predijera positivo a un único ejemplo (positivo) y el resto como negativo, tendría R muy baja, pero $P=1$.

Un clasificador perfecto tendría $P=R=1$.

Métrica de exactitud presenta el número de elementos correctamente clasificados en comparación con el número total de artículos, la métrica de exactitud no funciona bien con las clases que tienen muchos elementos de la misma clase (desquilibradas) (Ruiz Reina, 2018), para lo cual se presenta a continuación la ecuación propuesta por el autor en la ecuación (5).

$$ACURRACY = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Puntuación F, surge de la combinación de la métrica de precisión y la exhaustividad, donde la mejor puntuación $F1=1$ y la peor 0 (Fayrix, 2019), para lo cual se representa la ecuación propuesta por el autor en la ecuación (6).

$$F1 = \frac{2*PRECISION*RECALL}{PRECISION+RECALL} \quad (6)$$

7.3.13 DataSet

Es un conjunto de datos contenidos en una matriz de datos, la cual se constituye por filas y columnas, donde la columna representa una variable determinada y la fila el conjunto de datos del contexto en el que está trabajando. Al generar el conjunto de datos contenidos (Dataset) permite tener una mejor organización consultar y modificar el conjunto de datos comprendidos en la hoja de datos (Balagueró, 2018).

7.3.14 Excel

El software Excel, se le considera como una de las aplicaciones fundamentales para el desarrollo de actividades de cálculo, que permite trabajar con números, letras y gráficos de forma sencilla e intuitiva, dicho software cuenta con una interfaz dinámica y amigable que facilita su utilización, ya que al ingresar datos numéricos extensos permite realizar operaciones matemáticas de forma fácil (Ebriik, 2010).

8 HIPÓTESIS

Si se diseña un modelo para determinar factores de deserción, entonces se podrá obtener las causas y su nivel de influencia predictiva en la deserción de los estudiantes en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

Tabla 4: Descripción de Variables

VARIABLE	CONTENIDO DE LA PREGUNTA
Variable dependiente	Las causas y su nivel de influencia predictiva en la deserción de los estudiantes.
Variable independiente	Modelo para determinar factores de deserción.

Fuente: Grupo de trabajo

9 METODOLOGÍAS Y DISEÑO EXPERIMENTAL

9.1 Metodología

9.1.1 Metodología Científica

Esta metodología es un conjunto de procedimientos utilizada para adquirir conocimientos científicos, este método es riguroso y de orden lógico se puede comprobar la verdad o falsedad de los enunciados, además, este método es el camino para alcanzar a la meta. Son dos conceptos distintos el método son instrucciones que para lograr objetivos y la metodología es el estudio del método (Vásquez Hidalgo, 2005).

9.1.2 Tipos de investigación

9.1.2.1 Investigación Bibliográfica

Para el desarrollo del presente proyecto de investigación se aplicó la investigación bibliográfica, ya es la primera etapa del proceso investigativo que se involucra con el estudio sistemático de informes o escritos como fuentes de datos, tales como: tesis, revistas, libros y artículos científicos, puesto que esto proporciona un conocimiento mayor en el ámbito académico e investigativo. La presente investigación se constituye de técnicas y procedimientos que constituyen los medios instrumentales para su iniciación, teniendo como técnica principal la lectura pues a través de ella se adquiere el conocimiento (Méndez Rodríguez, 2008).

9.1.2.2 Investigación mixta

Para nuestro proyecto se utilizó la investigación mixta tanto cualitativa como cuantitativa, las mismas que permitieron conseguir una perspectiva más amplia sobre los factores de deserción en los estudiantes, con los cuales se obtuvo una mejor exploración y aprovechamiento de datos

que fueron analizados y a la vez alcanzando resultados satisfactorios, además estas técnicas mantienen sus propios procedimientos y estructuras (Pereira Pérez, 2011).

9.1.2.3 Investigación cuantitativa

Este tipo de investigación permitió conocer la realidad mediante la recopilación y análisis de los datos medibles, para ello se utilizó procedimientos que permite la medición de las mismas, con lo cual se logró obtener información de datos estadísticos y se definió las variables de mayor impacto en el proyecto de investigación de acuerdo a las respuestas obtenidas (Garcia, 2004).

9.1.2.4 Investigación cualitativa

Por medio de este tipo de investigación se estableció un análisis subjetivo, además se obtuvo datos que no son medibles basados en la entrevista, revisión de documentos, entre otros, con esto se estableció una reconstrucción de la realidad de la deserción de acuerdo a los criterios de los autores, con esta investigación se puede establecer una hipótesis (Portilla Chaves, Rojas Zapata, & Hernández Arteaga, 2014).

9.1.2.5 Investigación experimental

Para este proyecto se aplicó la investigación experimental, en la que se identificó las causas y efectos de la deserción estudiantil, permitió comprender como las variables influyen en los resultados, en este caso de las variables independientes que ocasionaron un efecto en la variable dependiente y a la vez se manipuló dichas variables de forma controlada para que los datos obtenidos sean representativos de la realidad (Arias, 2012).

9.1.3 Métodos de investigación

9.1.3.1 Método sintético

El presente método analiza y sintetiza la información recopilada, lo que permite ir estructurando las ideas, es decir, para el presente proyecto primero se recopila los datos, se analiza y se sintetiza la información adquirida. Una vez que se analiza dicha información se obtiene resultados sobre el objeto de estudio a fin de contribuir a la toma de decisiones, para evitar que los estudiantes deserten de sus estudios universitarios (Maya, 2014).

9.1.4 Técnicas de investigación

9.1.4.1 Encuesta

Para desarrollar el proyecto se ha considerado como mejor técnica de investigación la encuesta, la misma que se aplicó a los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas

de la Universidad Técnica de Cotopaxi, lo cual ayudó a obtener mediciones cuantitativas de un conjunto de características del tema de investigación. A partir de este procedimiento útil y sencillo, se pueden definir preguntas relevantes de acuerdo a las respuestas obtenidas de la población a las que se aplicó.

9.1.4.2 Entrevista

Se utilizó la entrevista como técnica de investigación para el desarrollo del proyecto en la que se interactuó con los directores de cada carrera de la Facultad de Ciencias y Aplicadas, en la que se obtuvo información sobre los factores de la deserción estudiantil. Luego que se realizó este método correspondiente y sin ningún inconveniente, se pudo comparar los distintos puntos de vista u opiniones de los entrevistados sobre el tema de deserción.

9.1.5 Instrumentos de investigación

9.1.5.1 Cuestionario

Se basa en un conjunto de preguntas sobre aspectos que se desean conocer de una investigación para obtener información que contribuyan al desarrollo del mismo (García Muñoz, 2014).

El cuestionario contiene preguntas cerradas que corresponden a si o no y de selección múltiple que contiene un listado de alternativas en las que el encuestado puede seleccionar una o varias opciones de acuerdo a su criterio.

9.1.5.2 Población

La población de estudio está constituida por los estudiantes de la Universidad Técnica de Cotopaxi sede la Matriz, para las carreras de: Sistemas de Información, Eléctrica, Electromecánica e Industrial pertenecientes a la Facultad de Ciencias de la Ingeniería y Aplicadas.

Tabla 5: Carreras de la Facultad de Ciencias de la Ingeniería y Aplicadas

Carreras	Número de estudiantes encuestados
Sistemas de información	449
Industrial	348
Electromecánica	297
Eléctrica	363
Total	1457

Fuente: Tomado de Departamento de TIC'S de la UTC (2019)

9.1.6 Métodos específicos

La metodología seleccionada para el modelo de minería de datos corresponde a la Knowledge Discovery in Databases (KDD), la misma que se refiere a un proceso de descubrir conocimiento útil almacenados en grandes bases de datos. Esta metodología consta de cinco fases que comprende: etapa de selección, etapa de preprocesamiento, etapa de transformación, etapa de extracción y finalmente la etapa de interpretación y evaluación.

La presente metodología se considera iterativa puesto que la salida de alguna de las fases logra hacer regresar a anteriores pasos, ya que a menudo son necesarias diversas iteraciones para extraer conocimiento de alta calidad, de igual manera es interactivo debido a que el usuario o experto en el dominio del problema debe ayudar en la preparación de datos y validar el conocimiento extraído entre otras actividades (Sarasa, Suárez, & Sánchez, 2008).

9.1.6.1 Etapa de selección

Para el autor (Timarán-Pereira, S. R., Hernández-Arteaga, Caicedo-Zambrano, S. J., Hidalgo-Troya, & Alvarado-Pérez, 2016) en la etapa de selección una vez identificado el conocimiento relevante, prioritario y definidas las metas del proceso Knowledge Discovery in Databases (KDD), desde el punto de vista del usuario final, se crea un conjunto de datos objetivo: seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La selección de los datos varía de acuerdo con los objetivos del negocio.

9.1.6.2 Etapa de pre-procesamiento

Según el autor (Timarán-Pereira, S. R., Hernández-Arteaga et al., 2016) en la etapa de pre-procesamiento/limpieza (data cleaning) se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos (noisy data), se seleccionan estrategias para el manejo de datos desconocidos (missing y empty), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista.

Los datos ruidos (noisy data) son valores significativos que se encuentran fuera del rango de valores esperados. Los datos desconocidos (empty) se refieren a aquellos a los cuales no les corresponde un valor en el mundo real y los missing, corresponden al valor que no fue capturado. Los datos nulos son datos desconocidos por los sistemas gestores de bases de datos relacionales (SGBDR). En la etapa de pre-procesamiento todos estos valores se ignoran, es

decir, se reemplazan por un valor más cercano, utilizando las métricas estadísticas como: la media, la moda, el mínimo y el máximo (Timarán-Pereira, S. R., Hernández-Arteaga et al., 2016).

9.1.6.3 Etapa de transformación

De acuerdo al autor (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) se buscan características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos.

Los métodos de reducción de dimensiones pueden simplificar una tabla de una base de datos horizontal o verticalmente. La reducción horizontal implica la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo por otro de alto nivel, en una jerarquía definida de valores categóricos o por la discretización de valores continuos (por ejemplo, edad por un rango de edades). La reducción vertical implica la eliminación de atributos que son insignificantes o redundantes con respecto al problema, como la eliminación de llaves, la eliminación de columnas que dependen funcionalmente (por ejemplo, edad y fecha de nacimiento). Se utilizan técnicas de reducción como agregaciones, compresión de datos, histogramas, segmentación, discretización basada en entropía, muestreo, entre otras (Berzal & Marín, 2016).

9.1.6.4 Etapa de extracción

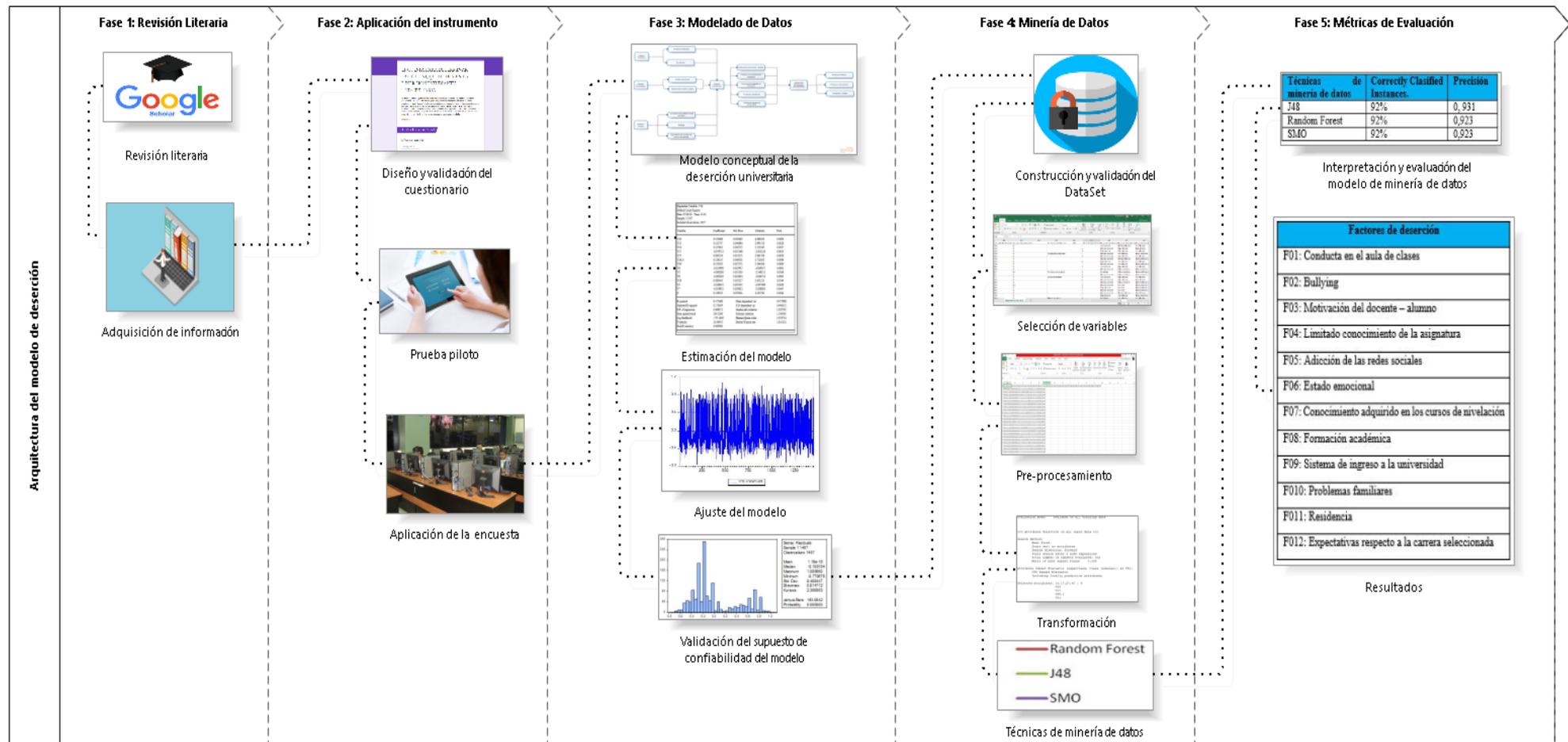
Según el autor (Timarán-Pereira, S. R., Hernández-Arteaga et al., 2016) el objetivo de la etapa de extracción es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación y la regresión. Las técnicas de minería de datos crean modelos que son predictivos o descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que se denominan variables dependientes, usando otras variables denominadas independientes o predictivas, como por ejemplo para determinar si los nuevos estudiantes desertan o no en función de su zona de procedencia, facultad, estrato, género, edad y promedio de notas. Los modelos descriptivos identifican patrones que explican o resumen los datos; sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos, como por ejemplo identificar patrones de compra de clientes en una determinada zona de la ciudad. Entre las tareas descriptivas se encuentran las reglas de asociación, los patrones secuenciales, los clustering y las correlaciones.

9.1.6.5 Etapa de interpretación y evaluación

De acuerdo al autor (Timarán-Pereira, S. R., Hernández-Arteaga et al., 2016) en la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para documentarlo y reportarlo a las partes interesadas; también la toma decisiones que permita verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

9.2 Diseño de la Investigación

Figura 10: Diseño de la investigación para determinar factores de deserción



Fuente: Grupo de trabajo

En la Figura 10, muestra el diseño de la investigación propuesta para determinar factores de deserción estudiantil. El modelo consta de cinco fases: en la primera fase se indagó sobre las concepciones teóricas de la deserción estudiantil en bases de datos científicas, en la segunda fase se construyó y validó el instrumento (cuestionario), la misma que fue aplicada a los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi, de este modo se obtuvo un DataSet que sirvió como punto de partida para el proceso de investigación. En la tercera fase se tiene la determinación del modelo para determinar factores de deserción estudiantil universitaria a través del software Eviews, en la cuarta fase se aplicó la metodología de minería de datos Knowledge Discovery in Databases (KDD), conformada por cinco etapas: selección, pre-procesamiento, transformación, extracción e interpretación y evaluación. Se emplearon cinco técnicas de minería de datos que son: Multilayer Perceptron, Naive Bayes, J48, Random Forest y Minimal Sequential Optimization (SMO). Las técnicas de minería de datos que arrojaron mayor exactitud fueron: J48, Random Forest y Minimal Sequential Optimization (SMO) con el 92%. Finalmente, en la fase cinco, se presenta los resultados en donde se contempla doce factores de deserción estudiantil que corresponden a: conducta en el aula de clases, bullying, motivación del docente – alumno, limitado conocimiento de la asignatura, adicción de las redes sociales, estado emocional, conocimiento adquirido en los cursos de nivelación, formación académica, sistema de ingreso a la universidad, problemas familiares, residencia y expectativas respecto a la carrera seleccionada.

10 ANÁLISIS Y DISCUSIÓN DE RESULTADOS

10.1 Recopilación de información

Se aplicó una encuesta online mediante Google Forms, para conocer la percepción del estudiante respecto a las causas que pueden ocasionar el abandono estudiantil. La encuesta consta de 59 preguntas entre cerradas y de selección múltiple, aplicadas en un periodo de un mes y medio, desde el 4 de abril hasta el 20 de mayo del 2019. La encuesta estuvo dirigida a los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas en las carreras de Electromecánica, Industrial, Eléctrica y Sistemas de Información de la Universidad Técnica de Cotopaxi matriculados en el periodo Marzo - Agosto 2019, en donde 1457 estudiantes dieron respuesta a la encuesta.

El objetivo de la encuesta fue descubrir los factores que afectan a la deserción estudiantil universitaria. La encuesta constó de tres secciones. La sección 1 recopiló información personal

y relacionada con la caracterización de la institución (26 preguntas), la sección 2 contenía preguntas que determinaron la influencia de los factores que inciden en la deserción (28 preguntas), y la sección 3 contenía preguntas que complementaron al estudio (5 preguntas).

Los resultados para la sección 2 fueron evaluados por dos indicadores donde Si = 1 y No = 0, otra alternativa fue a través de la escala de Likert (del 1 al 5), donde No influye = 1, Baja influencia = 2, Mediana influencia = 3, Alta influencia = 4, Influye totalmente = 5. Por otro lado se aplicó la escala de Likert que es un instrumento psicométrico donde el encuestado debe indicar su acuerdo o desacuerdo sobre una afirmación, ítem o reactivo, recopila datos cuantitativos de una investigación son cinco el número de opciones de respuestas más usado como se puede observar anteriormente, se tiene cinco ítems de evaluación para identificar factores de deserción estudiantil (Matas, 2018).

Para validar las preguntas de la encuesta, se estableció una prueba piloto, la cual fue dirigida a los estudiantes de noveno ciclo de la carrera de Sistemas de Información con un total de 42 personas encuestadas. El objetivo era verificar si las preguntas fueron entendibles, como resultado de esta prueba se corrigió las preguntas de la encuesta original utilizando un lenguaje fácil de entender, se recopiló información en el que se aplicó el proceso de minería de datos para obtener resultados.

Tabla 6: Resultados de la evaluación de la escala de Likert

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Baja Influencia	212	14,6	14,6	14,6
	Mediana Influencia	455	31,2	31,2	45,8
	Alta Influencia	587	40,3	40,3	86,1
	Influye Totalmente	203	13,9	13,9	100,0
	Total	1457	100,0	100,0	

En la Tabla 6 se muestra los resultados de la escala de Likert, la misma que consta de 16 variables, en un rango del 1 al 5 para medir las respuestas, se estableció de acuerdo al criterio del grupo de trabajo que 0 – 20 = 1, 21 – 40 = 2, 41 – 50 = 3, 51 – 60 = 4 y mayor a 61 = 5. Además, se obtuvo como resultado una frecuencia de 587 respuestas como Alta Influencia.

10.2 Confiabilidad de los datos

Se aplicó el alfa de Cronbach, para medir la fiabilidad de la encuesta, técnica utilizada para determinar el grado en que los elementos del cuestionario se relacionan entre sí. Como resultado se obtiene un valor de 0,839 a través del software estadístico Statistical Package for the Social Sciences (SPSS). Debido a que el valor obtenido supera el 0,7 se puede determinar que las preguntas de la encuesta aplicada son óptimas para el fin propuesto en la investigación. La tabla 7 presenta los resultados obtenidos.

Tabla 7: Resultados de validez de la encuesta

Estadísticos de fiabilidad	
Alfa de Cronbach	N de elementos
0,839	16

10.3 Análisis de la información

En la Tabla 8, se muestra un resumen de las características demográficas de los estudiantes matriculados en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

Tabla 8: Estadísticas descriptivas de Ciencias de Ingeniería y Aplicadas

Variable	Descripción	Valor	Porcentaje
Género	Masculino	1173	80.51%
	Femenino	284	19.49%
Edad	< 20	397	27.25%
	21-28	1026	70.42%
	>29	34	2.33%
Carrera	Sistemas de Información	449	30.82%
	Industrial	348	23.88%
	Eléctrica	363	24.91%
	Electromecánica	297	20.38%
Estado Civil	Casado	43	2.95%
	Divorciado	2	0.14%
	Soltero	1366	93.75%
	Unión Libre	45	3.09%
	Viudo	1	0.07%

Tabla 8: Estadísticas descriptivas de Ciencias de Ingeniería y Aplicadas (continuación)

Etnia	Mestizo	1380	94.71%
	Mulato	5	0.34%
	Blanco	10	0.69%
	Indígena	52	3.57%
	Afroecuatoriano	10	0.69%
Discapacidad	Si	14	0.96%
	No	1443	99.04%
Estatus Social	Bajo	278	19.08%
	Media	1168	80.16%
	Alto	11	0.75%
Vive cerca de la Universidad	Si	460	31.57%
	No	997	68.43%
Trabaja	Si	314	21.55%
	No	1143	78.45%
Tipo de Hogar	Nuclear Funcional (Padre, madre y hermanos)	1030	70.69%
	Materno Funcional (Madre y hermanos)	256	17.57%
	Nuclear Funcional (Cónyuge e hijos)	58	3.98%
	Inmediato Funcional (Otros Familiares)	113	7.76%
Tipo de Vivienda	Propia	1058	72.61%
	Arrendada	336	23.06%
	Hipotecada	16	1.10%
	Cedida	36	2.47%
	Otro	11	0.75%
Nivel educativo del padre	Ninguno	57	3.91%
	Centro alfabetización	24	1.65%
	Educación básica	689	47.29%
	Educación media	488	33.49%
	Superior universitaria completa	93	6.38%
	Superior universitaria incompleta	89	6.11%
	Posgrado maestría	17	1.17%

Tabla 8: Estadísticas descriptivas de Ciencias de Ingeniería y Aplicadas (continuación)

Nivel educativo de la madre	Ninguno	37	2.54%
	Centro alfabetización	63	4.32%
	Educación básica	692	47.49%
	Educación media	473	32.46%
	Superior universitaria completa	104	7.14%
	Superior universitaria incompleta	71	4.87%
	Posgrado maestría	17	1.17%
Colegio de Procedencia	Fiscal	1194	81.95%
	Municipal	31	2.13%
	Particular	160	10.98%
	Fiscomisional	65	4.46%
	Otro	7	0.48%

Fuente: Grupo de trabajo

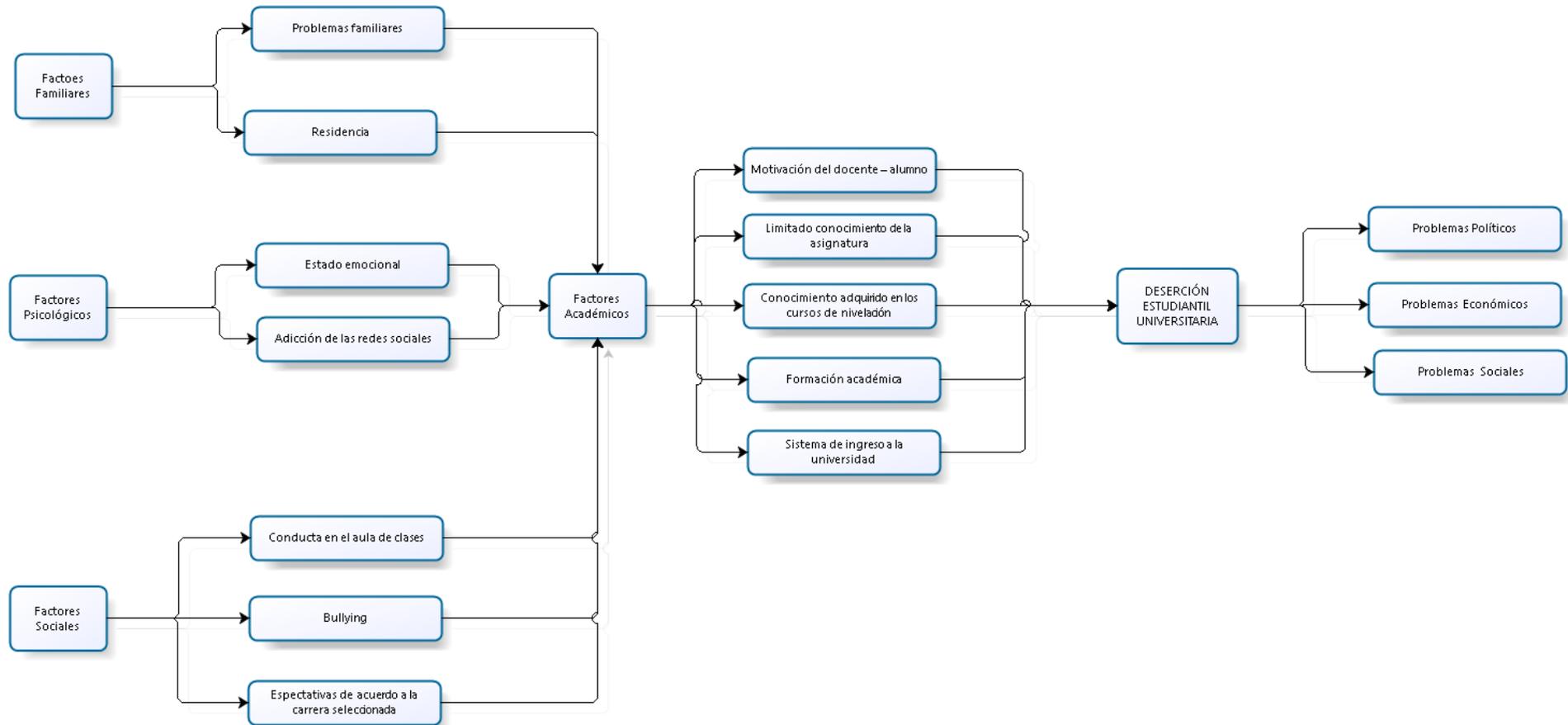
Los resultados presentados corresponden a los datos de los estudiantes encuestados, donde se puede evidenciar como datos prioritarios que la población masculina es más alta con 80.51% (1173 encuestados), mientras que la población femenina el 19.49% (284 encuestados), en los grupos comprendidos entre las edades de (21-28 años) se tiene 70.42% (1026 estudiantes), el 93.75% (1366 estudiantes) son solteros, el mayor porcentaje de la población encuestada se considera mestiza con el 94.71% (1380 estudiantes), el 99.04% (1443 estudiantes) no poseen ningún tipo de discapacidad, mientras tanto el 68.43% (997 estudiantes) viven lejos de la universidad, los estudiantes provienen de hogares conformada por padre, madre y hermanos el 70.69% (1030 estudiantes), generalmente residen en un domicilio propio el 72.61% (1058 estudiantes), el nivel educativo del padre es del 47.29% que corresponde a educación básica, mientras que el nivel de formación de la madre de los estudiantes en su mayoría es de educación básica con el 47.49% y un alto porcentaje de estudiantes que provienen de colegios fiscales con el 81.95% (1194 estudiantes).

10.4 Etapa de especificación del modelo

La Figura 11, presenta el modelo conceptual propuesto para determinar factores de deserción estudiantil. El modelo propuesto consta de cuatro dimensiones y doce factores que podrían ser considerados como causas de deserción en la universidad en la cual se aplicó el estudio.

10.4.1.1 Modelo para determinar factores de deserción estudiantil universitaria en la Facultad de Ciencias de la Ingeniería y Aplicadas.

Figura 11: Modelo conceptual



Fuente: Grupo de trabajo

10.4.1.2 Modelado de datos en Eviews

Regresión lineal

La regresión lineal simple consiste en una técnica estadística, la misma que analiza la relación entre dos variables cuantitativas para comprobar si la relación es lineal. Al tener más de dos variables se define como una regresión múltiple. Por otra parte, la finalidad de la regresión es explicar el comportamiento de una variable Y (dependiente o endógena) a partir de otra variable X (independiente o exógena) (Laguna, 2009).

Mínimos cuadrados ordinarios (MCO)

Es un método de estimación utilizada cuando se realiza un ajuste del modelo de regresión lineal simple al utilizar parámetros. Permite obtener un hiperplano para poder minimizar la suma de los cuadrados de las distancias entre las observaciones de la variable y el mismo hiperplano (residuos). Además, de que este método permite sumar las distancias al cuadro y define como otro supuesto razonable que es arbitrario cuando se trata de acumular las distancias que deben ser manejadas y entendidas (Chirivella Gonzáles 2019).

El método de ajuste por mínimos cuadrados o regresión lineal permite obtener la pendiente a de la recta y la ordenada en el origen, correspondientes a la recta $y = ax + b$ que mejor se ajusta a los n datos (x_i, y_i) , es decir, permite establecer una relación funcional entre dos variables; donde x es la variable independiente, (y) es la variable dependiente. En otras palabras (y) depende de (x) , en donde $(y) - (x)$ son dos variables importantes (Catarina 2006).

Según el autor (Hurtado Cruz, 2016) menciona que el método de mínimos cuadrados se aplica para ajustar rectas a una serie de datos presentados como punto en el plano.

Supongamos que se tienen los siguientes datos para las variables (x) , (y) .

Tabla 9: Variables (x, y)

X1	X2	Xn
Y1	Y2	Yn

Fuente: Tomado de Hurtado (2016)

Esta situación se puede presentar en estudios experimentales, donde se estudia la variación de cierta magnitud (x) en función de otra magnitud (y) .

Las expresiones de las variables sean de tipo lineal, según Hurtado en el 2016 se presenta en la ecuación (7).

$$y = mx + b \quad (7)$$

El método de mínimos cuadrados nos proporciona un criterio a través del cual se obtiene la mejor recta que representa a los puntos dados, las ecuaciones (8), (9), (10) se presentan en base al trabajo realizado por Hurtado (2016).

Se desearía tener

$$y_i = mx_i + b \quad (8)$$

Para todos los puntos (x_i, y_i) de $(i = 1), \dots, n$. Sin embargo, como general

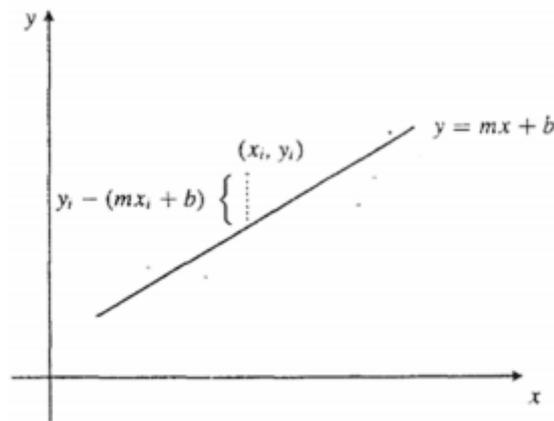
$$y_i \neq mx_i + b \quad (9)$$

Se pide que la suma de los cuadrados de las diferencias (las desviaciones)

$$y_i - mx_i + b \quad (10)$$

La Figura 12, presenta la gráfica de la regresión lineal y su ecuación correspondiente.

Figura 12: Ecuación generada



Fuente: Tomado de Hurtado (2016)

Expresiones para el estimador de mínimos cuadrados

Novales (2010) propone las ecuaciones de la regresión lineal a través de mínimos cuadrados los cuales se presentan en las ecuaciones (11), (12). En general, si primero despejamos β_0 en (11), tenemos:

$$\beta_0 = \frac{\sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i}{n} = \bar{y} - \beta_1 \bar{x} \quad (11)$$

Que se utiliza para estimar los mínimos cuadrados ordinarios (MCO) de β_0 , una vez que se tiene el estimador de 1. Sustituyendo en (12), se tiene:

$$\beta_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n S_{xy}}{n S_x^2} = r_{xy} \frac{S_y}{S_x} \quad (12)$$

10.4.1.3 Estimación del modelo

Se especifica el modelo a través de la técnica de regresión lineal para la predicción de los valores, así como también para el estudio de la relación entre variables cuantitativas. Para el caso de estudio la variable dependiente (Y) corresponde a: Deserción (VD) y 12 variables independientes o predictoras (X_1, X_2, \dots, X_n) que fueron conducta en el aula de clases (V10), bullying (V13), motivación del docente – alumno (V17), limitado conocimiento de la asignatura (V27), adicción de las redes sociales (V28.4), estado emocional (V30), conocimiento adquirido en los cursos de nivelación (V6), formación académica (V5), sistema de ingreso a la universidad (V8), problemas familiares (V19), residencia (V3), expectativas respecto a la carrera seleccionada (V7). Las cuales se representan en la Tabla 10.

Tabla 10: Descripción de los factores de deserción estudiantil

Variable independiente	Descripción
V10	Conducta en el aula de clases
V13	Bullying
V17	Motivación del docente – alumno
V27	Limitado conocimiento de la asignatura
V28.4	Adicción de las redes sociales
V30	Estado emocional
V6	Conocimiento adquirido en los cursos de nivelación
V5	Formación académica
V8	Sistema de ingreso a la universidad
V19	Problemas familiares
V3	Residencia
V7	Expectativas respecto a la carrera seleccionada
Variable dependiente	
VD	Deserción estudiantil universitaria.

Fuente: Grupo de trabajo

La estimación del modelo a través de mínimos cuadrados ordinarios utilizando el software Eviews, se presenta en la Tabla 11.

Tabla 11: Estimación del modelo inicial

Estimation Command: LS V16 V10 V13 V14 V17 V27 V28_4 V30 V6 V5 V8 C
Estimation Equation: $V16 = C(1)*V10 + C(2)*V13 + C(3)*V14 + C(4)*V17 + C(5)*V27 + C(6)*V28_4 + C(7)*V30 + C(8)*V6 + C(9)*V5 + C(10)*V8 + C(11)$
Substituted Coefficients: $V16 = 0.151695876998*V10 + 0.123390821825*V13 + 0.139288149971*V14 - 0.084202833337*V17 + 0.117401452718*V27 + 0.133946726279*V28_4 + 0.153413068987*V30 - 0.114032604774*V6 - 0.0963022550382*V5 - 0.091295282887*V8 + 0.166226211771$

El modelo propuesto representa una lista de serie en un formato tradicional de especificación, además una expresión algebraica con coeficientes simbólicos y de ecuaciones ya estimadas con el valor de los coeficientes resultantes, además se establece los principales factores de deserción y la determinación de los pesos resultantes en cada una de las variables.

Tabla 12: Modelo con todas variables

Dependent Variable: V16				
Method: Least Squares				
Date: 07/08/19 Time: 17:36				
Sample: 1 1457				
Included observations: 1457				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
V1	0.032459	0.029661	1.094324	0.2740
V10	0.139215	0.031259	4.453619	0.0000
V11	0.018534	0.022511	0.823315	0.4105
V12	0.005141	0.023352	0.220148	0.8258
V13	0.117354	0.041250	2.844928	0.0045
V14	0.132201	0.044321	2.982780	0.0029
V15	-0.047608	0.048409	-0.983458	0.3256
V17	-0.073588	0.028024	-2.625898	0.0087
V18	0.043947	0.025920	1.695517	0.0902
V19	0.055855	0.036883	1.514365	0.1302
V2	0.073370	0.111602	0.657421	0.5110
V20	0.028018	0.039314	0.712668	0.4762
V21	0.022772	0.042957	0.530110	0.5961

Tabla 12: Modelo con todas variables (continuación)

V22	-0.016494	0.037573	-0.438975	0.6607
V23	-0.031204	0.027691	-1.126833	0.2600
V24	-0.027759	0.024700	-1.123858	0.2613
V24_1	0.005154	0.034116	0.151086	0.8799
V25	-0.034382	0.026648	-1.290210	0.1972
V25_1	0.025495	0.042773	0.596050	0.5512
V25_2	0.033603	0.045120	0.744750	0.4565
V26_1	0.001540	0.027550	0.055897	0.9554
V27	0.076414	0.034897	2.189718	0.0287
V28	-0.011721	0.025007	-0.468711	0.6393
V28_1	-0.022296	0.027097	-0.822816	0.4108
V28_2	0.026358	0.055114	0.478241	0.6326
V28_3	-0.033736	0.024002	-1.405566	0.1601
V28_4	0.136795	0.048177	2.839448	0.0046
V28_5	-0.007937	0.029095	-0.272795	0.7851
V29	0.052860	0.034000	1.554696	0.1202
V3	-0.037152	0.023422	-1.586230	0.1129
V30	0.135748	0.029515	4.599328	0.0000
V32	0.030004	0.023920	1.254346	0.2099
V31	0.006672	0.024355	0.273936	0.7842
V4	-0.036073	0.024309	-1.483951	0.1380
V6	-0.101194	0.026315	-3.845516	0.0001
V5	-0.076722	0.032060	-2.393051	0.0168
V7	-0.051019	0.031628	-1.613124	0.1069
V8	-0.075719	0.023678	-3.197872	0.0014
V9	0.045715	0.030626	1.492678	0.1357
C	0.114051	0.087799	1.299006	0.1942
R-squared	0.193501	Mean dependent var		0.277968
Adjusted R-squared	0.171304	S.D. dependent var		0.448152
S.E. of regression	0.407965	Akaike info criterion		1.071798
Sum squared resid	235.8387	Schwarz criterion		1.216867
F-statistic	8.717350	Durbin-Watson stat		1.815898
Prob(F-statistic)	0.000000			

La Tabla 13, presenta el resultado del modelo ajustado después de eliminar las variables que estadísticamente no son significativas, lo cual se puede evidenciar por su peso correspondiente a valores inferiores a 0.05. Es importante resaltar que para la evaluación del modelo se consideró el 95 % de confianza para el indicador estadístico (Prob).

En la salida de la estimación se muestra los principales resultados obtenidos en relación con los parámetros de test de significancia y con el contraste de las hipótesis obtenidas, respecto a las 12 variables predictoras y el valor del coeficiente (β) que presenta el nivel de significancia estadística positiva y negativa respecto a cada variable identificada.

Tabla 13: Modelo ajustado de factores de deserción estudiantil

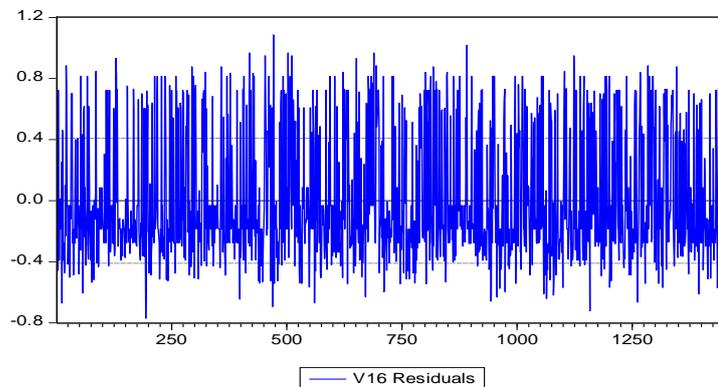
Dependent Variable: V16				
Method: Least Squares				
Date: 07/09/19 Time: 11:43				
Sample: 1 1457				
Included observations: 1457				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
V10	0.151689	0.030403	4.989343	0.0000
V13	0.121707	0.040684	2.991510	0.0028
V14	0.137861	0.043735	3.152165	0.0017
V17	-0.079711	0.027268	-2.923218	0.0035
V27	0.093539	0.031371	2.981709	0.0029
V28_4	0.126125	0.046321	2.722833	0.0066
V30	0.152303	0.027371	5.564504	0.0000
V6	-0.103990	0.025937	-4.009357	0.0001
V5	-0.080280	0.031503	-2.548311	0.0109
V8	-0.083203	0.022854	-3.640714	0.0003
V19	0.081445	0.033227	2.451211	0.0144
V3	-0.046455	0.023145	-2.007086	0.0449
V7	-0.059933	0.029832	-2.008983	0.0447
C	0.158533	0.070341	2.253791	0.0244
R-squared	0.177469	Mean dependent var		0.277968
Adjusted R-squared	0.170059	S.D. dependent var		0.448152
S.E. of regression	0.408271	Akaike info criterion		1.055791

Tabla 13: Modelo ajustado de factores de deserción estudiantil (continuación)

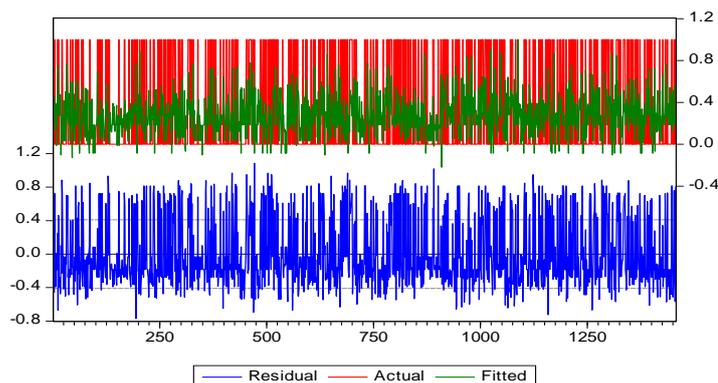
Sum squared resid	240.5268	Schwarz criterion	1.106565
Log likelihood	-755.1440	Hannan-Quinn criter.	1.074734
F-statistic	23.94933	Durbin-Watson stat	1.815132
Prob(F-statistic)	0.000000		

Una serie temporal es una secuencia de n observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o varias características (serie multivariante o vectorial) de una unidad observable en diferentes momentos (Marrón, 2008).

La Figura 13, presenta un nivel constante, la tasa de crecimiento se encuentra en orden cronológico; sin embargo, las desviaciones grandes (pequeñas) de dicha serie con respecto a su nivel van seguidas de variaciones grandes (pequeñas), por lo que su volatilidad condicionada por su historia reciente presenta algún tipo de inercia.

Figura 13: Residuos del modelo con la variable 16

La Figura 14, presenta una serie con tendencia común con un componente determinista, es decir, los mismos datos en orden aleatorio.

Figura 14: Continuidad del residuo del modelo

10.4.1.4 Etapa de pruebas

Supuesto de normalidad de los datos:

Es una prueba asintótica de muestras grandes, la misma que se basa en los mínimos cuadrados ordinarios (MCO), calculando como primera instancia la asimetría y la kurtosis de los residuos de los mínimos cuadrados ordinarios (MCO) (Geometry & Analysis, 2017).

Según Guajatira y Poter (2010), el estadístico de prueba corresponde a la ecuación (13).

$$JB = n * \frac{s^2}{6} + \frac{(K-3)^2}{24} \quad (13)$$

En donde:

n = representa el tamaño de la muestra

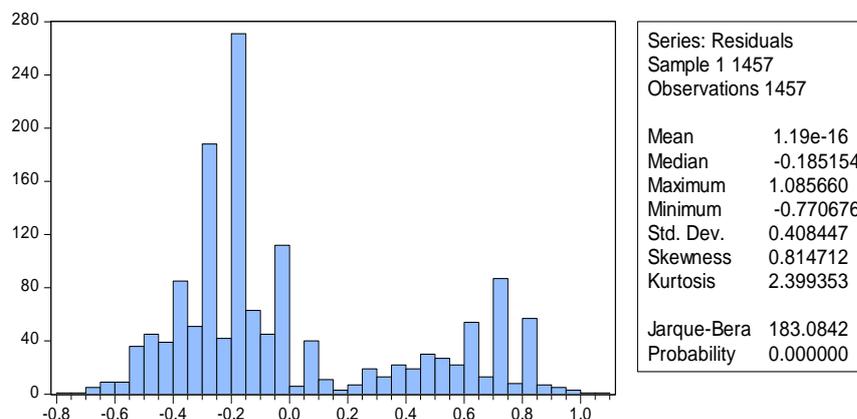
S = coeficiente de asimetría

K= coeficiente de kurtosis

Para la comprobación del supuesto de normalidad de los datos se utilizó la prueba estadística Jarque- Bera que consiste en la medida de bondad para los ajustes de desviación de la normalidad y está basada en los valores de kurtosis y asimetría de la muestra (Alfranca, 2009). La kurtosis es una métrica adimensional con el cual se determinara el grado de concentración que tienen los valores de la variable, en la zona central de la distribución de frecuencias (Quintero & Martín-landrove, 2019).

El indicador de probabilidad alcanzado a través de la técnica Jaque – Bera corresponde a 0.00 como se visualiza en la Figura 15, por lo que se puede señalar que existe normalidad en la distribución de los datos analizados ya que los términos de error están normalmente distribuidos.

Figura 15: Probabilidad Jarque-Bera



Supuesto de heteroscedasticidad

La heteroscedasticidad es aquel en que las varianzas de las perturbaciones no son constantes, por lo tanto, la variabilidad es diferente para cada observación. La matriz de varianzas-covarianzas es diagonal, por consiguiente, se sigue verificando independencia entre las observaciones aunque éstas no provienen de la misma población (Barreiro, Vicente, & Pintos Clapés, 2005).

A través de la prueba Breusch-Pagan, utilizada en la investigación para determinar la heteroscedasticidad en un modelo de regresión lineal. Analiza si la varianza estimada de los residuos de una regresión depende de los valores de las variables independientes (Barreiro et al., 2005).

Para determinar la existencia de heteroscedasticidad de los datos se aplican las siguientes hipótesis:

H_0 = No existe heteroscedasticidad en los datos.

H_1 = Que existe heteroscedasticidad en los datos.

Debido a que los resultados obtenidos se presentan en la Tabla 14, se acepta la hipótesis H_0 .

Tabla 14: Los resultados del proceso de la regresión lineal

Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	5.645624	Prob. F(2,1444)	0.0036	
Obs*R-squared	11.30451	Prob. Chi-Square(2)	0.0035	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 07/08/19 Time: 17:48				
Sample: 1 1457				
Included observations: 1457				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
V10	-0.000114	0.030390	-0.003744	0.9970
V13	0.000443	0.040699	0.010878	0.9913
V14	-0.003606	0.043792	-0.082353	0.9344
V17	-0.001186	0.027056	-0.043828	0.9650
V27	-0.003646	0.030322	-0.120241	0.9043

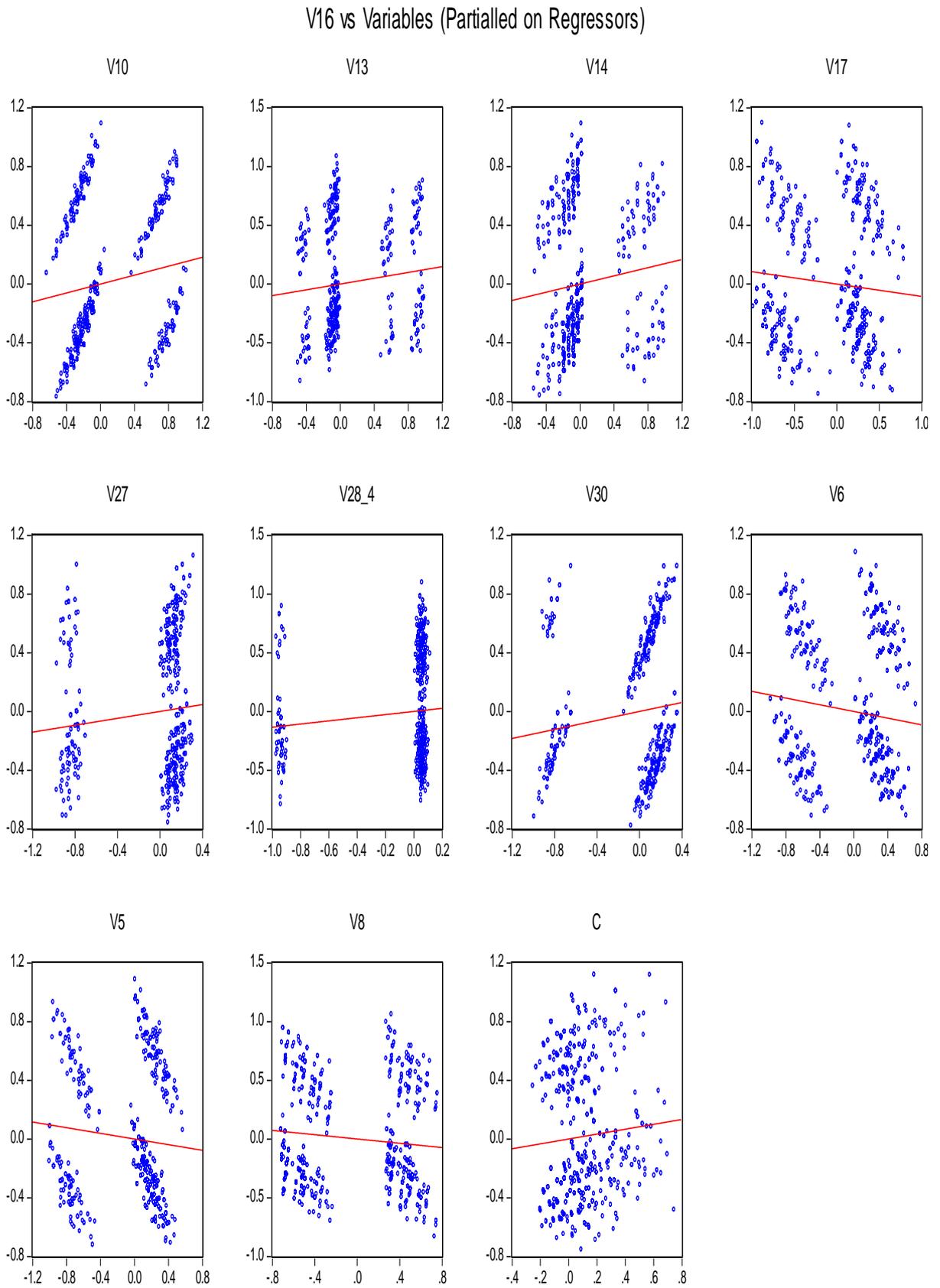
Tabla 14: Los resultados del proceso de la regresión lineal (continuación)

V28_4	0.000586	0.046284	0.012651	0.9899
V30	-0.001096	0.027336	-0.040084	0.9680
V6	-0.001046	0.025763	-0.040591	0.9676
V5	5.12E-05	0.030782	0.001665	0.9987
V8	-0.000426	0.022376	-0.019038	0.9848
C	0.005643	0.065949	0.085562	0.9318
RESID(-1)	0.086092	0.026378	3.263761	0.0011
RESID(-2)	0.013716	0.026450	0.518568	0.6041
R-squared	0.007759	Mean dependent var		1.19E-16
Adjusted R-squared	-0.000487	S.D. dependent var		0.408447
S.E. of regression	0.408546	Akaike info criterion		1.056459
Sum squared resid	241.0182	Schwarz criterion		1.103607
Log likelihood	-756.6307	Hannan-Quinn criter.		1.074049
F-statistic	0.940937	Durbin-Watson stat		1.997478
Prob(F-statistic)	0.504562			

La prueba de heteroscedasticidad (cuando los errores no son constantes a lo largo de toda la muestra) de Breusch-Pagan-Godfrey acepta la hipótesis nula de homoscedasticidad en el nivel de significación del 1%, la misma que permitió la evaluación de validez de las suposiciones del modelado.

Mínimo Cuadrados Ordinarios

La Figura 16, representa el resultado del método de mínimos cuadrados ordinarios respecto a las variables resultantes residuales como factores influyentes positivos y negativos con la deserción universitaria.

Figura 16: Regresión lineal

Conducta en el aula de clases (V10): presenta una relación directamente proporcional con la deserción. Este factor se puede explicar debido a que la conducta en el aula de clases es considerada como un problema, cuando un estudiante emite una conducta agresiva porque reacciona ante un conflicto. La relación de la conducta agresiva del estudiante puede originarse de problemas de relación social con otros estudiantes o docentes. Una conducta agresiva del estudiante podría ser considerada como una práctica negativa que ha conducido a las malas relaciones interpersonales que genera malestar social. La manera de comportarse de una persona puede ser considerada como un factor de deserción (Freire Bayas, 2014).

Bullying(V13): podría ser considerado como una causante de la deserción, este problema relaciona signos como burlas, insultos o una exclusión social que supone indiferencia entre los estudiantes dentro de las instituciones educativas (Lladó & Mares, 2017). Presenta afectaciones en el desempeño académico causando la deserción (Millan Reyes, Barrera Sánchez, & Ospina Diaz, 2015).

Motivación del docente – alumno (V17): se refiere al incentivo emocional recibido por el cumplimiento adecuado con tareas asignadas. De igual manera puede ser considerada como el conjunto de principios (atención, respeto, cordialidad, responsabilidad, entre otros) de enseñanza y educación (García Rangel, García Rangel, & Reyes Angulo, 2014). La motivación docente-alumno puede ser identificada como factor que influye en la deserción estudiantil debido a la falta de interés del alumno con su formación académica (García Rangel et al., 2014). Al contrario, un estudiante motivado, puede alcanzar sus metas de éxito académico. Por tal razón se considera importante que la motivación es un estudio fundamental a la hora de analizar la deserción.

Limitado conocimiento de la asignatura (V27): se convierte en un factor causante de la deserción relacionada a la situación del estudiante en el desarrollo de cada asignatura y el desempeño en las mismas (Ariza & Marín, 2009). Las deficientes bases académicas, la dificultad en el empleo de estrategias de aprendizaje incide a la pérdida de ciertas asignaturas provocando que el estudiante no se sienta tranquilo, conforme y seguro con lo que realiza (Ariza & Marín, 2009).

Adicción de las redes sociales (V28.4): se conoce como aquella conducta repetida que resulta placentera y genera una pérdida de control en el sujeto con una interferencia grave en su vida cotidiana, a nivel familiar, laboral o social. Así, en las redes sociales se puede acceder a contenidos sin fines de educación. Las personas adictas a las redes sociales experimentan un

síndrome de abstinencia cuando no pueden conectarse, caracterizado por la presencia de un profundo malestar emocional (estado de ánimo disfórico, insomnio, irritabilidad e inquietud psicomotriz) (Echeburúa Odriozola, 2012). Podría considerarse una causante de la deserción ya que en un estudio realizado en la comunidad estudiantil de Madrid, encontraron que uno de cada tres estudiantes de entre los que mantienen más de un perfil en redes sociales no tienen un buen rendimiento académico puesto que le dedican poco tiempo al estudio (Tamez Aguirre, 2012).

Estado emocional (V30): se define como alteraciones en el desarrollo cognitivo y psicomotor del estudiante, sin que en general puedan ser asignados a categorías diagnósticas específicas tales como retardo mental, síndrome de déficit atencional o trastornos específicos del aprendizaje. Las dificultades emocionales constituyen un serio problema tanto para la educación y para los mismos estudiantes, provocando el bajo rendimiento y por ende la deserción estudiantil (Jadue, 2002).

Conocimiento adquirido en los cursos de nivelación (V6): se conoce como un espacio de colaboración para paliar los efectos negativos de la calidad de la educación secundaria recibida por estos estudiantes, través del cual se podrían actualizar planes y programas de estudio, para así poder proporcionar a los alumnos bases sólidas para cursar el nivel superior (Frites & Miranda, 2014). En base a lo mencionado se puede manifestar que afecta a la deserción ya que produce un número creciente de alumnos reprobados y desertores por bajos conocimientos (Cu Balán, 2008).

Formación académica (V5): Una inadecuada formación académica puede contribuir a la deserción significa que la universidad, por medio de sus programas y acciones pedagógicas concretas, propone una educación en valores y en responsabilidad social, lo que se refleja en el conjunto de conocimientos adquiridos por los estudiantes (Aragón Tobón, Clavijo Zapata, Puerta Lopera, & Sánchez Duque, 2014). La deserción de los estudiantes del sistema formal evidencia la existencia de situaciones de riesgo que llevan a los alumnos a no continuar su formación académica en un centro educativo (Echeburúa Odriozola, 2012).

Sistema de ingreso a la universidad (V8): el acceso a la Universidad en el Ecuador requiere de un examen de ingreso, el cual es una prueba tanto de conocimiento como de aptitud académica a fin de obtener información sobre las habilidades y destrezas con las que cuentan los estudiantes, mismas que serán indispensables para el postulante y le asegurarán el éxito, dentro del sistema educativo de pregrado (Vieira, 2017). En base al resultado de las encuestas aplicadas

se puede identificar que el 41% de los estudiantes indica que ellos no eligieron la carrera por vocación, así como también la institución en la que estudian, provocando que el fenómeno de la deserción aumente, conforme a la variación en el sistema de ingreso a carrera.

Problemas familiares (V9): son problemas que se suscitan en el hogar, por ejemplo: la violencia, el desinterés y la falta de preocupación por parte de los padres para con sus hijos, problemas de salud, entre otros (Peña, Soto, & Calderón, 2016). Los jóvenes que, agobiados por problemas familiares de diversa índole, corren el riesgo de abandonar sus estudios, debido a que estos problemas pueden ocasionar un excesivo porcentaje de faltas permitidas en la reglamentación universitaria (Zavala Soto, 2017).

Residencia (V3): el lugar de residencia es un factor importante debido a los efectos relevantes que tiene sobre la deserción, ya que está relacionada con el lugar donde vive el estudiante (Osorio, Bolancé, & Castillo Caicedo, 2019). Se considera un factor en el caso de estudio, debido a que el 42% de los alumnos que contestaron la encuesta residen en las provincias de Tungurahua y Pichincha. Esto puede ocasionar atrasos al inicio de sus actividades académicas, gasto en el transporte diario, bajas calificaciones, entre otros.

Expectativas carrera (V7): con respecto a la carrera seleccionada, la deficiente orientación vocacional recibida antes de ingresar a estudiar una carrera profesional, provoca que los estudiantes se inscriban en las carreras sin sustentar su decisión en una sólida información sobre las mismas. Lo que provoca poco interés por los estudios en general afectando a la deserción (González Fiegehen, 2006). Se puede evidenciar debido a que el 41% de la población encuestada aseguran no estar estudiando la carrera que ellos desean.

10.4.2 Análisis de datos a través de técnicas de minería de datos

10.4.2.1 Fase 1: Integración y limpieza de los datos

Como resultado de la aplicación de la encuesta a los estudiantes se obtuvo 42 factores de deserción. A través del proceso estadístico mediante el software Eviews se pudo identificar 12 factores que estadísticamente se consideran como significativas, estos factores representan en su coeficiente (β) de signos positivos o negativos que presentan el nivel de influencia en la deserción.

10.4.2.2 Fase 2: Preprocesamiento (selección de variables)

Comparación de los métodos en Waikato Environment for Knowledge Analysis (WEKA)

Una de las opciones que posee WEKA es la pestaña de selección de atributos, cuya función consiste en identificar mediante un conjunto de datos ciertos atributos. El peso asociado a aquellos atributos o valores característicos que determinan si los datos son de una clase u otra (García Morate, 2008). Para el presente caso de estudio, se realizó la comparación entre el método: BestFirst y el atributo: CfsSubsetEval.

La función del método BestFirst, consiste en una colina codiciosa escalada con retroceso, que permite ver la correlación que existe entre las diferentes variables, mientras que el atributo CfsSubsetEval, considera el valor predictivo de cada atributo individualmente, junto con el grado de redundancia entre ellos (Gutiérrez García, 2016). El resultado de la aplicación de este método se presenta en la Tabla 15.

Tabla 15: Método: BestFirst - Atributo: CfsSubsetEval

Attribute Selection on all input data
Search Method: Best first. Start set: no attributes Search direction: forward Stale search after 5 node expansions Total number of subsets evaluated: 302 Merit of best subset found: 0.055
Attribute Subset Evaluator (supervised, Class (nominal): 42 VD): CFS Subset Evaluator Including locally predictive attributes
Selected attributes: 16,17,27,40 : 4 V16 V17 V25.1 V31

Luego se aplicó el método GreedyStepwise, que se refiere a la escalada de colina codiciosa sin retroceso; opcionalmente genera una lista clasificada de atributos como se puede visualizar en la Tabla 16.

Tabla 16: Método: GreedyStepwise - Atributo: CfsSubsetEval

Attribute Selection on all input data
Search Method: Greedy Stepwise (forwards). Start set: no attributes Merit of best subset found: 0.055
Attribute Subset Evaluator (supervised, Class (nominal): 42 VD): CFS Subset Evaluator Including locally predictive attributes
Selected attributes: 16,17,27,40 : 4 V16 V17 V25.1 V31

Por medio del software WEKA se realizó una de la técnica CfsSubsetEval a través de los métodos BestFirst y GreedyStepwise. La Tabla 15, 16 presenta el resultado obtenido del proceso de selección de variables en la cual se puede visualizar los pasos asignados para cada variable.

En la Tabla 17, se visualiza los doce factores de deserción estudiantil con su respectiva descripción y código.

Tabla 17: Factores de deserción

Factor	Descripción	Código
Conducta en el aula de clases	Comprende el problema que se produce cuando un estudiante emite una conducta agresiva porque reacciona ante un conflicto. La relación de la conducta agresiva del estudiante puede originarse de problemas de relación social con otros estudiantes o docentes (Freire Bayas, 2014).	V10
Bullying	Son signos como burlas, insultos o una exclusión social que supone indiferencia entre los estudiantes dentro de las instituciones educativas (Lladó & Mares, 2017).	V13
Motivación del docente – alumno	Se refiere al incentivo emocional recibido por el cumplimiento adecuado con tareas asignadas (García Rangel et al., 2014).	V17
Limitado conocimiento de la asignatura	Se relaciona a la situación del estudiante en el desarrollo de cada asignatura y el desempeño en las mismas (Ariza & Marín, 2009).	V27
Adicción de las redes sociales	Se conoce como aquella conducta repetida que resulta placentera y genera una pérdida de control en el sujeto con una interferencia grave en su vida cotidiana, a nivel familiar, laboral o social (Echeburúa Odriozola, 2012).	V28.4

Tabla 17: Factores de deserción

Estado emocional	Corresponde a alteraciones en su desarrollo cognitivo y psicomotor del estudiante, sin que en general puedan ser asignados a categorías diagnósticas específicas tales como retardo mental, síndrome de déficit atencional o trastornos específicos del aprendizaje (Jadue, 2002).	V30
Conocimiento adquirido en los cursos de nivelación	Se conoce como un espacio de colaboración para paliar los efectos negativos de la calidad de la educación secundaria recibida por estos estudiantes, través del cual se podrían actualizar planes y programas de estudio, para así poder proporcionar a los alumnos bases sólidas para cursar el nivel superior (Frites & Miranda, 2014).	V6
Formación académica	Significa que la universidad, por medio de sus programas y acciones pedagógicas concretas, propone una educación en valores y en responsabilidad social, lo que se refleja en el conjunto de conocimientos adquiridos por los estudiantes (Aragón Tobón et al., 2014).	V5
Sistema de ingreso a la universidad	Se refiere a una prueba tanto de conocimiento como de aptitud académica a fin de obtener información sobre las habilidades y destrezas con las que cuentan los estudiantes, mismas que serán indispensables para el postulante y le asegurarán el éxito, dentro del sistema educativo de pregrado (Vieira, 2017).	V8
Problemas familiares	Problemas que se suscitan en el hogar por ejemplo: la violencia, el desinterés y la falta de preocupación por parte de los padres para con sus hijos, problemas de salud, entre otros (Peña, Soto, & Calderón, 2016).	V19
Residencia	Lugar donde reside el estudiante (Osorio et al., 2019).	V3
Expectativas respecto a la carrera seleccionada	La deficiente orientación vocacional recibida antes de ingresar a estudiar una carrera profesional, provoca que los estudiantes se inscriban en las carreras sin sustentar su decisión en una sólida información sobre las mismas (González Fiegehen, 2006).	V7

Fuente: Grupo de Trabajo

10.4.2.3 Fase 3: Extracción

En la fase de extracción para determinar factores que inciden en la deserción estudiantil universitaria, se utilizó la herramienta Weka en la que se determinaron cinco técnicas las cuales fueron: Multilayer Perceptron, Naive Bayes, J48, Random Forest y Sequential Minimal Optimization (SMO).

Conceptualización de las técnicas

Multilayer Perceptron: actúa como un aproximador universal de funciones, para la aplicación en las redes neuronales, se trabaja el algoritmo backpropagation, que significa un algoritmo de

aprendizaje supervisado, es decir muestra al menos una capa oculta con suficientes números de unidades no lineales. Esta red tiene la capacidad de generalización, facilidad de dar salidas satisfactorias a entradas que el sistema no ha visto en la fase de entrenamiento (Marín Diazaraque, 2012).

Matemáticamente el algoritmo backpropagation, según el autor (Vidal González, 2014), en el aprendizaje de la red se formula como un problema de minimización del siguiente modo:

$$\text{Min}WE \quad (14)$$

Siendo W: el conjunto de parámetros de la red, pesos y umbrales.

E: una función error que evalúa la diferencia entre las salidas de la red y las salidas deseadas.

La función error generalmente se define como:

$$E = \frac{1}{N} \sum_{n=1}^N e(n) \quad (15)$$

Donde N: es el número de patrones o muestras y e(n): es el error cometido por la red, para el patrón n, dado por:

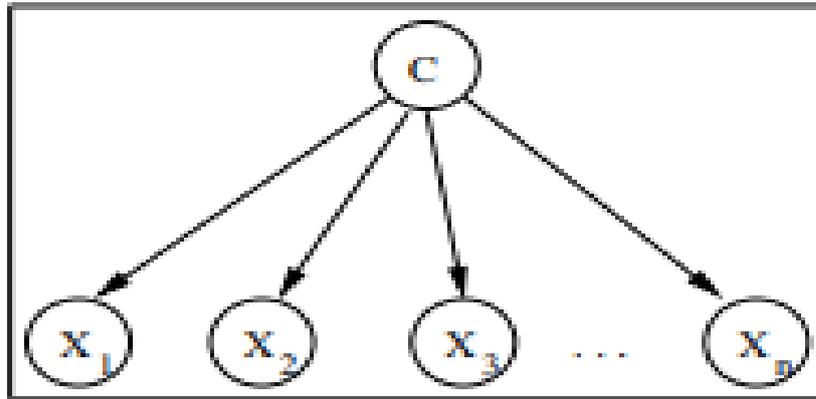
$$e(n) = \frac{1}{nC} \sum_{i=1}^C (si, (n) - yi(n))^2 \quad (16)$$

Naive Bayes: combina la clasificación y predicción con la finalidad de construir modelos para predecir posibles resultados a partir de la asociación de datos históricos (Dueñas Reyes, 2009). Otros autores consideran que es un problema de decisión binario con dos posibles decisiones (d_1 frente a d_2), una función de decisión es una función continua (Larrañaga, Inza, & Moujahid, 2010).

La función matemática comprende:

$$r = R^2 \rightarrow R \quad (17)$$

Figura 17: Estructura gráfica de un modelo Naive Bayes



Fuente: Tomado de Larrañaga, Inza & Moujahid (2018)

Verificando que si $r(x) > 0$ ($r(x) < 0$) se prefiere d_1 a d_2 (d_2 a d_1).

La función $r(x) = 0$ define una superficie de decisión.

J48: creado por Ross Quinlan en 1984, el cual es conocido también como un algoritmo de inducción que genera una estructura de árbol a partir de subconjuntos o ventanas extraídas de un grupo de datos de entrenamiento, además se considera que el árbol de decisión J48, tiene una alta precisión en comparación con los clasificadores SMO y Naive Bayes (Vizcaino Garzón, 2008).

El funcionamiento del algoritmo J48, consiste en los nodos internos que son los diferentes atributos, las ramas son los posibles valores y nodos finales (hojas), que corresponden a la clasificación, por ello se considera un método iterativo debido a que va colocando los posibles valores de las características (información ganada). Cuando todas caigan en una clasificación y ya no exista ambigüedad entonces se asigna una raíz o un nodo. Un árbol consta de una raíz y puede tener dos o más nodos hijos intermedios. En la mayoría de los árboles se considera un nodo intermedio como un nodo de predicción, la cual nos dirá según el valor numérico porque rama debe bajar. Esto puede ser visto como nodos circulares para los de predicción y rectangulares para los de decisión (Condori Castro & Sulla Torres, 2016).

Random Forest: se caracteriza por correr eficientemente sobre grandes bases de datos, proporciona el método experimental para detectar interacciones entre variables, esto proporciona las estimaciones para saber que variables son importantes en la clasificación (Escobar., Burbano., & Puris., 2016).

Random Forest, es un clasificador que consiste en una colección de estructura de árbol clasificador:

$$\{h(x, k), k = 1, \dots\} \quad (18)$$

Donde $\{k\}$ se distribuyen de manera idéntica e independiente. Los vectores aleatorios y cada árbol emiten un voto unitario para la clase más popular en la entrada x (Breiman, 2012).

Optimización Mínima Secuencial: la máquina de soporte vectorial (SVM), corresponde al conjunto de algoritmos de aprendizaje supervisado desarrollado por Vladimir Vapnik. En la implementación de la máquina de soporte vectorial (SVM) que utiliza weka, se encuentra la optimización mínima secuencial (SMO), la misma que para la clasificación considera cada atributo como una variable, así se obtiene varios atributos. El objetivo de la optimización mínima secuencial (SMO), es encontrar el hiperplano, es decir el vector que tiene una dimensión menos, que divide a los ejemplos en dos conjuntos y maximiza la distancia entre los puntos más cercanos entre los dos conjuntos (Zaballa Pardo, 2016).

Según el autor (Zaballa Pardo, 2016) debido a la restricción de igualdad lineal que involucra los multiplicadores de Lagrange α_1 , el problema más pequeño posible involucra dos de estos multiplicadores. Entonces, para cualesquiera dos multiplicadores. α_1 y α_2 , las restricciones se reducen a:

$$0 \leq \alpha_1, \alpha_2 < C, \quad (19)$$

$$y_1 \alpha_1 + y_2 \alpha_2 = k, \quad (20)$$

K: negativo de la suma sobre el resto de términos en la restricción de igualdad, que se fija en cada iteración.

En base a la revisión literaria, se considera que las técnicas antes mencionadas son importantes ya que aportan al proceso de predicción en nuestro trabajo de investigación.

Resultados

Para la aplicación del método Multilayer Perceptron establece un listado de especificaciones, dichos valores son predefinidos, se seleccionó el BatchSize de 100 números de instancias para procesar un lote de predicción. Para el porcentaje de entrenamiento se empleó el 66% y un 44% del conjunto de datos para la prueba. Se utilizó la validación cruzada (cross-validation) a través de 10 iteraciones, este método divide el conjunto total en subconjuntos de tamaño 10 partes. Se presenta las especificaciones del método Multilayer Perceptrón en la Tabla 18.

Tabla 18: Especificaciones del Multilayer Perceptrón

Especificaciones	VD = Variable dependiente	Deserción estudiantil universitaria
	Variables independientes	El DataSet posee 39 variables independientes
	Entrenamiento	66%
	Prueba	44%
	Validación cruzada	10
	GUI	False
	AutoBuild	True
	BatchSize	100
	Debug	False
	Decay	False
	DoNotCheckCapabilities	False
	HiddenLayers	a
	LearningRate	0.3
	Momentum	0.2
	NominalToBinaryFilter	True
	NormalizeAttributes	True
	NormalizeNumericClass	True
	NumDecimalPlaces	2
	Reset	True
	Seed	0
TrainingTime	500	
ValidationSetSize	0	
ValidationThreshold	20	
Resultados	Variables independientes	V16, V17, V25.1, V27, V31 y V40
	Número de instancias correctamente clasificadas	1311 que corresponde al 89%

Fuente: Grupo de trabajo

En la Tabla 19, se presenta la especificación del modelo de red neuronal a través del método Multilayer Perceptrón, se obtuvo una clasificación correcta de instancias con una tasa del 89% de precisión de la predicción. La función Sigmoid Node, se lo utiliza para la propagación hacia atrás y datos asociados. Existen nodos en la capa oculta de esta red, tal es el caso de los nodos de salida que son unidades lineales, la función Sigmoid Node 0 representa la unidad de salida y los Sigmoid Node 1- 22 son las unidades ocultas, estos valores están generados por los pesos de interconexión.

Tabla 19: Resultado de la herramienta que utiliza Weka que es Multilayer Perceptron

```

==== Run information ====
Scheme:   weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Relation: FactSeleccionados1-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 1457
Attributes: 42

```

V1	V15	V26
V2	V16	V26.1
V3	V17	V27
V4	V19	V28
V5	V20	V28.1
V6	V21	V28.2
V7	V22	V28.3
V8	V23	V28.4
V9	V24	V28.5
V10	V24.1	V29
V11	V25	V30
V12	V25.1	V31
V13	V25.2	V32
V14	V15	VD

```

Test mode: 10-fold cross-validation
==== Classifier model (full training set) ====
Sigmoid Node 0
  Inputs  Weights
  Threshold 7.3389939192401235
  Node 2 -5.7419680153361
  Node 3 -2.3970778428181894
  Node 4 -2.66464822000334
  Node 5 4.439584268815969
  Node 6 4.47672230286076
  Node 7 -2.9135484276038577
  Node 8 4.4348181453143
  Node 9 -5.439707819375495
  Node 10 -3.4454912736852474
  Node 11 3.517813685668383
  Node 12 -6.1525124384760215
  Node 13 3.4308423284398026
  Node 14 4.03057919159796
  Node 15 -4.139941274603706
  Node 16 -5.638672202631695
  Node 17 3.4906257269055145
  Node 18 -5.373443556532639
  Node 19 2.958748385734836
  Node 20 -1.9244030471189448
  Node 21 -0.8028404281829332
  Node 22 -2.7767038395194845
Sigmoid Node 1
  Inputs  Weights
  Threshold -7.337988148335526
  Node 2 5.739259631581039
  Node 3 2.3966674075701895
  Node 4 2.6642605888475117
  Node 5 -4.44046476057559
Sigmoid Node 2
  Inputs  Weights
  Threshold -0.3657309965578198
  Attrib V1=1 -1.1272419852151072
  Attrib V2=1 -0.12545353829055533
  Attrib V3=1 -0.9421038573552217
  Attrib V4=1 3.493614236660276
  Attrib V5=1 -1.357656266054015
  Attrib V6=1 -1.1651981942626302
  Attrib V7=1 1.7560055257910268
  Attrib V8=1 -0.5856958875573436
  Attrib V9=1 -0.20881141896998845
  Attrib V10=1 0.5522275893987788
  Attrib V11=1 -1.2885152134933373
  Attrib V12=1 -2.4073036439877473
  Attrib V13=1 -1.670182613662965
  Attrib V14=1 -0.7079103722376588
  Attrib V15=1 1.1729807294318106
  Attrib V16 =1 -0.8989606653119488
  Attrib V17=1 -0.8737406361042998
  Attrib V18=1 -0.5576346450620406
  Attrib V19=1 -0.21465178759407563
  Attrib V20=1 0.1829949788087961
  Attrib V21=1 -0.5973960622374795
  Attrib V22=1 -0.00734438638380642
  Attrib V23=1 2.9447907863814766
  Attrib V24=1 -0.07898205796317677
  Attrib V24.1=1 -0.26838011160160735
  Attrib V25=1 0.5347999880250028
  Attrib V25.1=1 -1.150043459566163
  Attrib V25.2=1 -0.12403679263017521

```

Naive Bayes esta conformada por especificaciones que permiten definir algunas características para los valores, así con el NumDecimalPlaces define el número de decimales de salida y el DisplayModelInOldFormat que usa un formato antiguo para la salida del modelo. Se utiliza el formato antiguo cuando hay muchos valores de clase y el formato nuevo cuando hay menos clases. Así también como el BatchSize de 100 números de instancias para procesar un lote de predicción. Para el porcentaje de entrenamiento se empleó el 66% y un 44% del conjunto de datos para prueba. Además, el valor de la validación cruzada (cross-validation) fue 10 iteraciones, es decir divide el conjunto total en subconjuntos de tamaño 10 partes. Se presenta las especificaciones de la técnica Naive Bayes en la Tabla 20.

Tabla 20: Especificaciones del Navie Bayes

Especificaciones	VD = Variable dependiente	Deserción estudiantil universitaria
	Variables independientes	El DataSet posee 39 variables independientes
	Entrenamiento	66%
	Prueba	44%
	Validación cruzada	10
	BatchSize	100
	Debug	False
	DisplayModelInOldFormat	False
	DoNotCheckCapabilities	False
	NumDecimalPlaces	2
	UseKernelEstimator	False
	UseSupervisedDiscretization	False
Resultados	Variables independientes	V16, V17, V25.1, V27, V31 y V40
	Número de instancias correctamente clasificadas	1283 que corresponde al 88%

Fuente: Grupo de trabajo

En la Tabla 21, se muestra el resultado obtenido con la técnica Naive Bayes, el tiempo que se demoró en ejecutar fue de 0.03 segundos. Las instancias correctamente clasificadas se denominan exactitud que corresponden al 88% que representa 1283 encuestas. Para la obtención de los resultados se empleó el DataSet compuesto por las 42 variables independientes (predictoras), además arroja los resultados de 1 y 0 puesto que se trabajó con datos binarios, seguidamente muestra el resultado total por variables.

Tabla 21: Resultado de la herramienta que utiliza Weka que es Naive Bayes

```

==== Run information ====
Scheme:   weka.classifiers.bayes.NaiveBayes
Relation: FactSeleccionados1-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 1457
Attributes: 42

```

V1	V15	V26
V2	V16	V26.1
V3	V17	V27
V4	V19	V28
V5	V20	V28.1
V6	V21	V28.2
V7	V22	V28.3
V8	V23	V28.4
V9	V24	V28.5
V10	V24.1	V29
V11	V25	V30
V12	V25.1	V31
V13	V25.2	V32
V14	V15	VD

Time taken to build model: 0.03 seconds

```

==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances   1283      88.0577 %
Incorrectly Classified Instances  174      11.9423 %
Kappa statistic                  0.1852
Mean absolute error              0.1517
Root mean squared error         0.3041
Relative absolute error         106.4705 %
Root relative squared error     114.1586 %
Total Number of Instances       1457

```

Test mode: 10-fold cross-validation								
==== Classifier model (full training set) ====								
Naive Bayes Classifier								
Attribute	Class 0	1	Attribute	Class 0	1	Attribute	Class 0	1
	(0.08)	(0.92)		(0.08)	(0.92)		(0.08)	(0.92)
V1			V15			V26		
0	23.0	263.0	0	15.0	69.0	0	48.0	546.0
1	91.0	1084.0	1	99.0	1278.0	1	66.0	801.0
[total]	114.0	1347.0	[total]	114.0	1347.0	[total]	114.0	1347.0
V2			V16			V26.1		
0	112.0	1333.0	0	55.0	999.0	0	40.0	426.0
1	2.0	14.0	1	59.0	348.0	1	74.0	921.0
[total]	114.0	1347.0	[total]	114.0	1347.0	[total]	114.0	1347.0
V3			V17			V27		
0	75.0	924.0	0	62.0	272.0	0	23.0	198.0
1	39.0	423.0	1	52.0	1075.0	1	91.0	1149.0
[total]	114.0	1347.0	[total]	114.0	1347.0	[total]	114.0	1347.0

En la Tabla 22, se describe la salida de pantalla que arrojó la herramienta Weka con el árbol de decisión J48, el tiempo de ejecución fue de 0.12 segundos. Así mismo, se obtuvo una clasificación correcta de instancias que se denominana exactitud con un valor de 92% de precisión de la predicción. El árbol de decisión esta conformado por 13 hojas y con un tamaño del árbol de 25, siendo la Motivación del docente- alumno (V17) el nodo raíz o padre, además de sus nodos intermedios e hijos, estos nodos son de predicción y de decisión. En la Figura 18, se presenta el árbol de decisión generado a través de la herramienta Weka.

Figura 18: Árbol de decisión

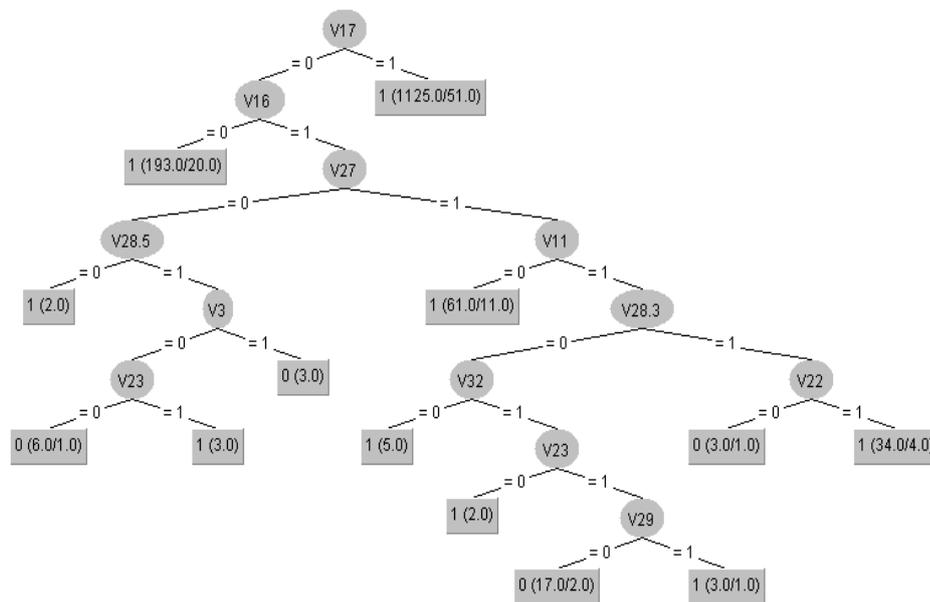


Tabla 22: Resultado de la herramienta que utiliza Weka que es J48

```

==== Run information ====
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: FactSeleccionados1-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 1457
Attributes: 42

          V1          V15          V26
          V2          V16          V26.1
          V3          V17          V27
          V4          V19          V28
          V5          V20          V28.1
          V6          V21          V28.2
          V7          V22          V28.3
          V8          V23          V28.4
          V9          V24          V28.5
          V10         V24.1         V29
          V11         V25          V30
          V12         V25.1         V31
          V13         V25.2         V32
          V14         V15          VD

Test mode: 10-fold cross-validation

```

Tabla 22: Resultado de la herramienta que utiliza Weka que es J48 (continuación)

```

==== Classifier model (full training set) ====
J48 pruned tree
-----
V17 = 0
| V16 = 0: 1 (193.0/20.0)
| V16 = 1
| | V27 = 0
| | | V28.5 = 0: 1 (2.0)
| | | V28.5 = 1
| | | | V3 = 0
| | | | V23 = 0: 0 (6.0/1.0)
| | | | V23 = 1: 1 (3.0)
| | | | V3 = 1: 0 (3.0)
| | | V27 = 1
| | | | V11 = 0: 1 (61.0/11.0)
| | | | V11 = 1
| | | | | V28.3 = 0
| | | | | V32 = 0: 1 (5.0)
| | | | | V32 = 1
| | | | | | V23 = 0: 1 (2.0)
| | | | | | V23 = 1
| | | | | | | V29 = 0: 0 (17.0/2.0)
| | | | | | | V29 = 1: 1 (3.0/1.0)
| | | | | V28.3 = 1
| | | | | V22 = 0: 0 (3.0/1.0)
| | | | | V22 = 1: 1 (34.0/4.0)
V17 = 1: 1 (1125.0/51.0)

Number of Leaves:      13
Size of the tree:     25
Time taken to build model: 0.12 seconds
==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances   1341      92.0384 %
Incorrectly Classified Instances  116      7.9616 %
Kappa statistic                 0.1557
Mean absolute error              0.1294
Root mean squared error          0.2697
Relative absolute error          90.8442 %
Root relative squared error      101.2563 %
Total, Number of Instances      1457

```

El árbol de decisión J48 contiene especificaciones, tal es el caso del `confidenceFactor` que es un factor de confianza para la moda, al igual que el `minNumObj` que define el número mínimo de instancias por hoja. Al igual que en el `BatchSize` de 100 números de instancias para procesar un lote de predicción. Para el porcentaje de entrenamiento se empleó el 66% y un 44% del conjunto de datos para la prueba. Además, el valor de la validación cruzada (cross-validation) fue 10 iteraciones, es decir divide el conjunto total en subconjuntos de tamaño 10 partes. Se describe la especificación del árbol de decisión J48 en la Tabla 23.

Tabla 23: Especificaciones del J48

Especificaciones	VD = Variable dependiente	Deserción estudiantil universitaria
	Variables independientes	El DataSet posee 39 variables independientes
	Entrenamiento	66%
	Prueba	44%
	Validación cruzada	10
	BatchSize	100
	BinarySplits	False
	CollapseTree	True
	ConfidenceFactor	0.25
	Debug	False
	DoNotCheckCapabilities	False
	DoNotMAkeSplitPointActua IValue	False
	MinNumObj	2
	NumDecimalPlaces	2
	NumFolds	3
	ReduceErrorPruning	False
	SaveInstanceData	False
	Seed	1
	SubtreeRaising	True
Unpruned	False	
UseLaplace	False	
UseMDLcorrection	True	
Resultados	Variables independientes	V16, V17, V25.1, V27, V31 y V40
	Número de instancias correctamente clasificadas	1341 que corresponde al 92%

Fuente: Grupo de trabajo

En la Tabla 24, se describe el resultado generado con el árbol de decisión RandomForest, el mismo que tuvo un tiempo de ejecución de 0.5 segundos, las intancias correctamente clasificadas se denominana excatitud correspondiente a un 92% de precisión de la predicción. Dicho árbol de decisión clasifica los árboles mediante bosques el cual busca el camino más cercano para la predicción, se trabajó con un empaquetado de 100 iteraciones y base de aprendizaje.

Tabla 24: Resultado de la herramienta que utiliza Weka que es Random Forest

```

==== Run information ====
Scheme:   weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation: FactSeleccionados1-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 1457
Attributes: 42
           V1           V15           V26
           V2           V16           V26.1
           V3           V17           V27
           V4           V19           V28
           V5           V20           V28.1
           V6           V21           V28.2
           V7           V22           V28.3
           V8           V23           V28.4
           V9           V24           V28.5
           V10          V24.1         V29
           V11          V25           V30
           V12          V25.1         V31
           V13          V25.2         V32
           V14          V15           VD

Test mode: 10-fold cross-validation
==== Classifier model (full training set) ====
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.5 seconds
==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances   1343           92.1757 %
Incorrectly Classified Instances  114            7.8243 %
Kappa statistic                 -0.0027
Mean absolute error             0.1339
Root mean squared error         0.2576
Relative absolute error         94.0082 %
Root relative squared error     96.6819 %
Total, Number of Instances     1457

```

En la Tabla 25, el árbol de decisión RandomForest, esta conformada por especificaciones verdaderas y falsas respectivamente, entre una de las características se define el MaxDepth que define la profundidad del árbol, utilizando 0 para ilimitado. Además, el BatchSize de 100 números de instancias para procesar un lote de predicción. Para el porcentaje de entrenamiento se empleó el 66% y un 44% del conjunto de datos para la prueba. Además, el valor de la validación cruzada (cross-validation) fue 10 iteraciones, es decir divide el conjunto total en subconjuntos de tamaño 10 partes.

Tabla 25: Especificaciones del Random Forest

Especificaciones	VD = Variable dependiente	Deserción estudiantil universitaria
	Variables independientes	El DataSet posee 39 variables independientes
	Entrenamiento	66%
	Prueba	44%
	Validación cruzada	10
	BagSizePercent	100
	BatchSize	100
	BreakTiesRandomly	False
	CalcOutOfBag	False
	ComputeAttributeImportance	False
	Debug	False
	DoNotCheckCapabilities	False
	MaxDepth	0
	NumDecimalPlaces	2
	NumExecutionSlots	1
	NumFeatures	0
	NumIterations	100
	OutputOutOfBagComplexityStatistics	False
	PrintClassifiers	False
	Seed	1
StoreOutOfBagPredictions	False	
Resultados	Variables independientes	V16, V17, V25.1, V27, V31 y V40
	Número de instancias correctamente clasificadas	1343 que corresponde al 92%

Fuente: Grupo de trabajo

La presente Tabla 26, contiene el resultado con el Minimum Sequential Optimization (SMO), el tiempo que se demoró en ejecutar fue de 0.55 segundos. Las instancias correctamente clasificadas se denominan exactitud que corresponden al 92% que representa 1345 encuestas. El resultado contiene la Máquina lineal que muestra los pesos de los atributos, no los vectores de soporte.

Tabla 26: Resultado de la herramienta que utiliza Weka que es SMO

```

==== Run information ====
Scheme:          weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator
"weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
Relation:        FactSeleccionados1-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances:       1457
Attributes:      42

                V1          V15          V26
                V2          V16          V26.1
                V3          V17          V27
                V4          V19          V28
                V5          V20          V28.1
                V6          V21          V28.2
                V7          V22          V28.3
                V8          V23          V28.4
                V9          V24          V28.5
                V10         V24.1        V29
                V11         V25          V30
                V12         V25.1        V31
                V13         V25.2        V32
                V14         V15          VD

==== Classifier model (full training set) ====
SMO
Kernel used:
  Linear Kernel:  $K(x,y) = \langle x,y \rangle$ 
Classifier for classes: 0, 1
BinarySMO
Machine linear: showing attribute weights, not support vectors.
+ -0 * (normalized) V27=1
+ -0 * (normalized) V28=1
+ -0.0001 * (normalized) V28.1=1
+ 0.0006 * (normalized) V28.2=1
+ 0.0002 * (normalized) V28.3=1
+ -0.0001 * (normalized) V28.4=1
+ 0.0002 * (normalized) V28.5=1
+ -0.0001 * (normalized) V29=1
+ -0.0001 * (normalized) V30=1
+ -0.0001 * (normalized) V31=1
+ 0 * (normalized) V32=1
+ 1.0003
Number of kernel evaluations: 858294 (72.121% cached)
Time taken to build model: 0.55 seconds
==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances   1345      92.313 %
Incorrectly Classified Instances  112      7.687 %
Kappa statistic                  0
Mean absolute error              0.0769
Root mean squared error          0.2773
Relative absolute error          53.9557 %
Root relative squared error      104.0788 %
Total, Number of Instances      1457

```

El Minimum Sequential Optimization (SMO), esta definido por especificaciones, el Epsilon que define el error de redondeo, al igual que el ToleranceParameter que es el parámetro de tolerancia que no se debe cambiar. Así como, el BatchSize de 100 números de instancias para procesar un lote de predicción. Para el porcentaje de entrenamiento se empleó el 66% y un 44% del conjunto de datos para la prueba. Además, el valor de la validación cruzada (cross-validation) fue 10 iteraciones, es decir divide el conjunto total en subconjuntos de tamaño 10 partes. Se presenta la especificación de la técnica SMO en la Tabla 27.

Tabla 27: Especificaciones del SMO

Especificaciones	VD = Variable dependiente	Deserción estudiantil universitaria
	Variables independientes	El DataSet posee 39 variables independientes
	Entrenamiento	66%
	Prueba	44%
	Validación cruzada	10
	BatchSize	100
	BuildCalibrationModels	False
	C	1.0
	Calibrator	Logistic: -R1.0E-8-M-1 -num-decimal-places 4
	ChecksTurnedOff	False
	Debug	False
	DoNotCheckCapabilities	False
	Epsilon	1.0E-12
	FilterType	Normalize training data
	Kernel	Polykernel -E1.0-C250007
	NumDecimalPlaces	2
	NumFolds	-1
RandomSeed	1	
ToleranceParameter	0.001	
Resultados	Variables independientes	V16, V17, V25.1, V27, V31 y V40
	Número de instancias correctamente clasificadas	1345 que corresponde al 92%

Fuente: Grupo de trabajo

10.4.2.4 Fase 4: Interpretación y evaluación

Para la etapa de interpretación y evaluación se aplicaron métricas de evaluación que corresponden a: la Exactitud, Kappa Statistic, F-Measure, ROC Area y Precisión.

Tabla 28: Instancias clasificadas por técnicas de minería de datos

Técnicas de minería de datos	Correctly Classified Instances.
Multilayer Perceptron	89%
Naive Bayes	88%
J48	92%
Random Forest	92%
SMO	92%

Fuente: Grupo de trabajo

Al emplear las técnicas de minería de datos que posee la herramienta Weka, se obtuvo como resultado que J48, Random Forest y SMO, tiene más alto porcentaje de instancias correctamente clasificadas. Con esta tasa de precisión de la predicción del 92%, lo que indica que el modelo puede ser considerado como óptimo en términos de confiabilidad.

Por otro lado, para la evaluación del modelo basado en minería de datos para determinar factores de deserción estudiantil universitaria, se realizó a través de métricas de evaluación como es la precisión y la confiabilidad, el nivel del modelo propuesto se determinó en términos de confianza y fiabilidad del modelo de predicción.

Como se puede evidenciar en la Tabla 29, los coeficientes corresponden a valores de la ROC Area cercanos a 1 especialmente en la técnica de Naive Bayes por lo tanto se comprueba que el modelo propuesto podría ser considerado como óptimo en técnicas de predicción de la deserción.

Tabla 29: Resultados de las técnicas de minería de datos que posee Weka

Algoritmos	Kappa Statistic	Mean Absolute Error	F -Measure	ROC Área
Multilayer Perceptron	0.14	0.10	0.947	0.673
Naive Bayes	0.18	0.15	0.935	0.742
J48	0.15	0.12	0.958	0.610
Random forest	-0.002	0.13	0.959	0.716
SMO	0	0.07	0.960	0.500

Fuente: Grupo de trabajo

ThresholdCurve

Según el autor (García Morate, 2008) La visualización de la curva de umbral, muestra la variación de las proporciones de cada clase. Para nuestro modelo, se situó en el eje (x) los coeficientes falsos positivos y en el eje (y) los coeficientes de verdaderos positivos y de esa manera se obtuvo el área bajo la curva (ROC), por ende, aparecerá sobre la gráfica el valor del área de ésta. En la Figura 19, se presenta el resultado del proceso experimental aplicando las técnicas de minería de datos que posee la herramienta Weka que es Naive Bayes, en la Figura 20: el árbol de decisión J48, en la Figura 21: el Minimal Sequential Optimization (SMO), en la Figura 22: el árbol de decisión Random Forest y en la Figura 23: la red neuronal Multilayer Perceptron, se puede evidenciar el valor encontrado de la curva ROC, por cada una de las técnicas antes mencionadas.

Figura 19: ThresholdCurve con 1 - Naive Bayes

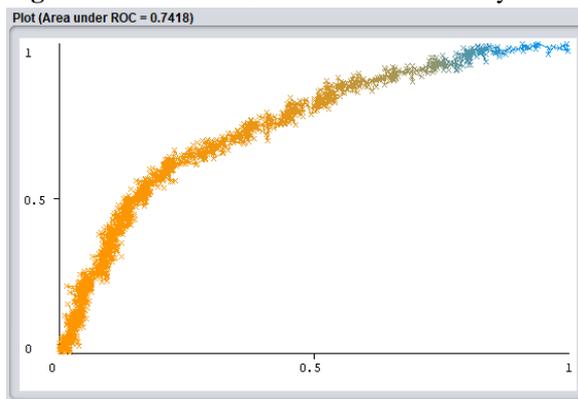


Figura 20: ThresholdCurve con 1 - J48

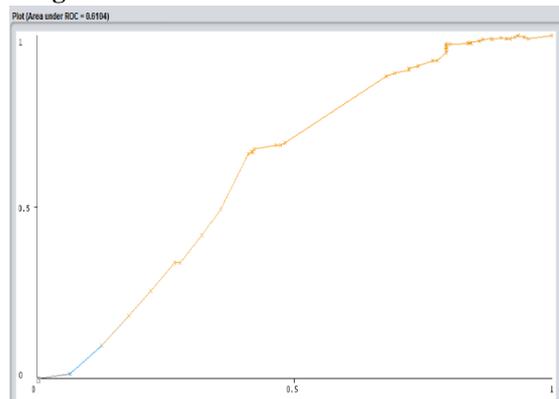


Figura 21: ThresholdCurve con 1 - SMO

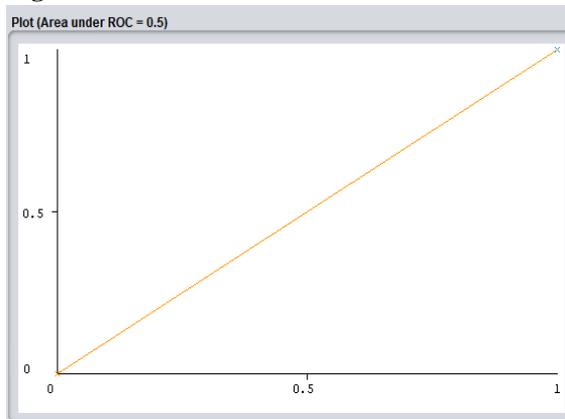


Figura 22: ThresholdCurve con 1 - RandomForest

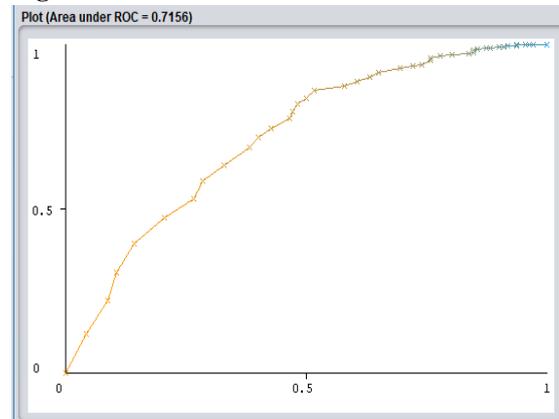
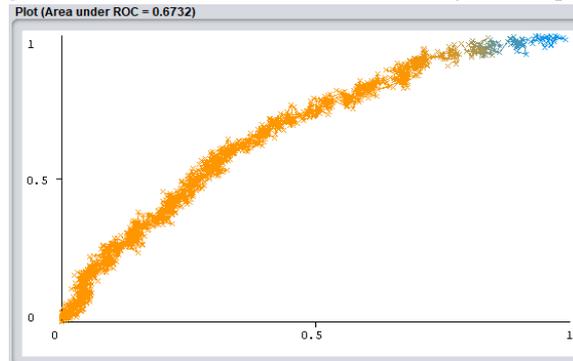


Figura 23: ThresholdCurve con 1 – Multilayer Perceptrón

CostCurve

Según el autor (García Morate, 2008) el costo de la curva, también conocido como visualize cost curve, muestra una gráfica que indica la probabilidad de coste a la sensibilidad entre clases. De igual manera, se presenta en la Figura número 24 el CostCurve del Naive Bayes, en la Figura 25 el J48, en la Figura 26 el Minimal Sequential Optimization (SMO), en la Figura 27 el Random Forest y en la Figura 28 el Multilayer Perceptron, haciendo referencia a cada técnica de minería de datos aplicados en el proceso de determinación de factores que inciden en la deserción estudiantil universitaria.

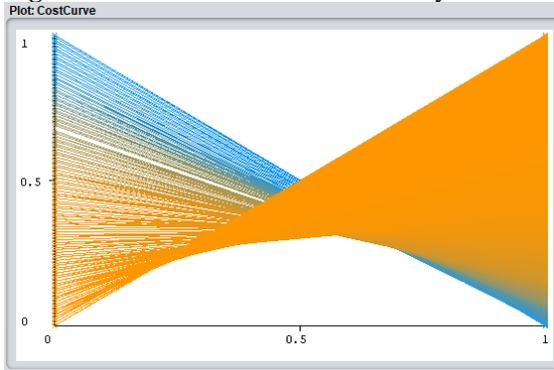
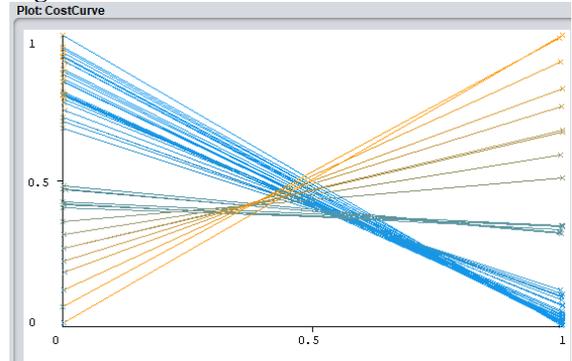
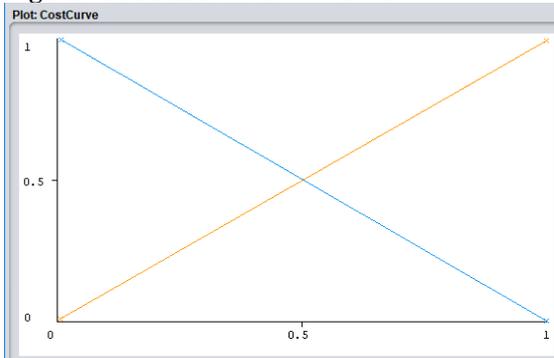
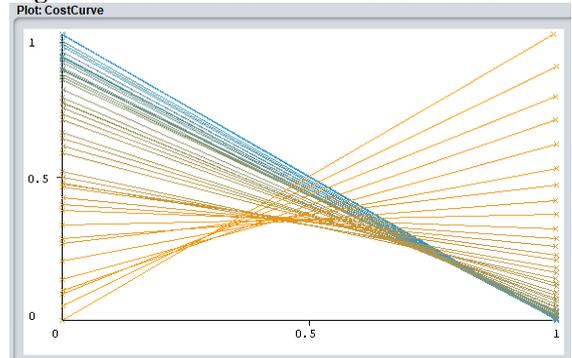
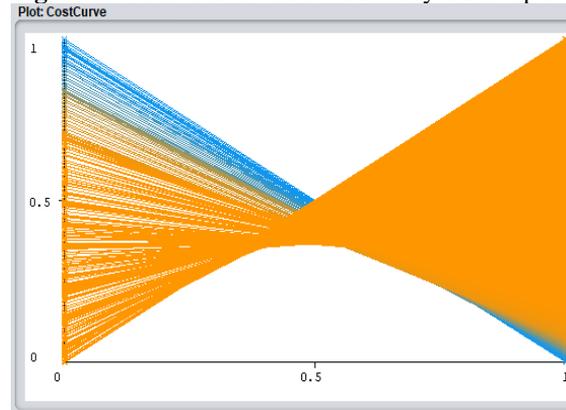
Figura 24: CostCurve con 1 - Naive Bayes**Figura 25:** CostCurve con 1 - J48**Figura 26:** CostCurve con 1 – SMO**Figura 27:** CostCurve con 1 – RandomForest

Figura 28: CostCurve con 1 – Multilayer Perceptrón

11 IMPACTOS (INSTITUCIONALES, SOCIALES, AMBIENTALES O ECONÓMICOS)

11.1 Impacto institucional

Con el modelo basado en minería de datos para determinar factores que inciden en la deserción estudiantil universitaria, se puede implementar estrategias adecuadas que permitan la reducción de las tasas de deserción en la Facultad de Ciencias de la Ingeniería y Aplicadas y a la vez contribuyan a mitigar la deserción, empleando medidas correctivas y oportunas para garantizar la estabilidad estudiantil.

11.2 Impacto económico

Se considera uno de los impactos más importantes ya que permitirá conocer el efecto que genera la deserción estudiantil a nivel económico, permite tener una eficiencia interna de la institución en situaciones como la reducción del gasto administrativo. Además, el incremento de retención de los estudiantes permitirá un aumento del presupuesto institucional universitario.

11.3 Impacto social

Beneficia a los estudiantes porque se podría identificar los problemas que les atañe a través de un seguimiento a los mismos y por medio de programas de retención, se puede contribuir a que los estudiantes culminen con éxitos sus estudios y por ende obtengan un título profesional.

Una vez que el estudiante concluya con su formación académica puede insertarse en el campo laboral, tener una situación económica estable que permita mejorar su condición de vida, disminuir las tasas de desempleo y lo más importante contribuir al desarrollo del país.

12 PRESUPUESTO PARA LA PROPUESTA DEL PROYECTO

En este apartado se muestran los gastos directos e indirectos que se generaron en el proyecto de investigación, así como los gastos aproximados a continuación se detalla cada uno de los gastos mencionados.

12.1 Gastos directos

Tabla 30: Descripción gastos directos

DETALLE	CANTIDAD	V. UNITARIO	V. TOTAL
Impresiones	500	0,07	35,00
Copias	30	0,05	1,50
Cuaderno	1	1,00	1,00
Esfero	2	0,35	0,70
Lápiz	1	0,50	0,50
Flash Memory	1	7,00	7,00
Internet	300hrs	0,60	180,00
Anillados	4	7,00	28,00
Empastado de Tesis	1	15,00	15,00
TOTAL			\$268,70

Fuente: Grupo de trabajo

12.2 Gastos indirectos

Tabla 31: Descripción gastos indirectos

DETALLE	CANTIDAD	PERSONAS	V. UNITARIO	V. TOTAL
Transporte a la universidad	5 meses	1	20,00	100,00
	5 meses	1	12,00	60,00
Alimentación	5 meses	2	20,00	100,00
TOTAL				\$260,00

Fuente: Grupo de trabajo

12.3 Gastos aproximados

Tabla 32: Total, Gastos Directos e Indirectos

Descripción	Total
Gastos Directos	\$268,70
Gastos Indirectos	\$260,00
Imprevistos +10%	\$52,87
Total, de Gastos	\$581,57

Fuente: Grupo de trabajo

13 CONCLUSIONES Y RECOMENDACIONES

13.1 Conclusiones

Este apartado del documento se presenta las conclusiones obtenidas del proceso de investigación:

- La revisión de la literatura permitió la sustentación bibliográfica de 100 documentos, entre tesis de grado y artículos científicos en fuentes de datos de investigación científica, como resultado se obtuvo la determinación del marco teórico, así como también se seleccionó la metodología de minería de datos a emplear, de igual manera las herramientas informáticas que permitieron el análisis de los datos y la correspondiente determinación de los factores de deserción estudiantil.
- Se obtuvo un DataSet elaborado a partir de la encuesta de 1457 estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas, los datos recopilados contribuyeron a generar el conjunto de datos (DataSet) que sirvió como punto de partida para el proceso de investigación, donde se obtuvieron las variables del modelo conceptual.
- Se diseñó un modelo conceptual de los factores de deserción estudiantil, enfocado en doce variables obtenidas a través de la encuesta aplicada a los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas, las variables resultantes en este modelo se convierten en doce hipótesis probadas a través de procesos experimentales que permiten evidenciar la influencia de estos factores en la deserción.
- Se validó el modelo por medio de métricas de evaluación, a través de las cuales se obtuvo que: el J48, Random Forest y SMO corresponden al 92% de las instancias correctamente clasificadas, respecto a otras técnicas aplicadas.

13.2 Recomendaciones

- Como trabajo futuro, se recomienda aplicar el mismo proyecto de investigación en las demás Facultades de la Universidad Técnica de Cotopaxi, pero en este caso tomando en cuenta sólo a los estudiantes de primero a quinto semestre, puesto que, en la revisión de la literatura, éstos son los ciclos más propensos en riesgo de desertar a fin de determinar los factores por los cuales los estudiantes de las otras facultades desertan de sus estudios.
- La implementación de un sistema informático para la determinación de estudiantes en riesgo de desertar.

- Identificación de los estudiantes que presentan uno o varios factores de deserción estudiantil previamente determinados durante la investigación.
- Para este caso de estudio se pudo realizar una construcción del modelo mediante la utilización de técnicas de minería de datos o inteligencia artificial como la Red Neuronal de Aprendizaje Profundo debido a que esta técnica no pudo ser utilizada, ya que no se contaba con el software de licenciamiento correspondiente para realizar los procesos experimentales.

14 BIBLIOGRAFÍA

- Alania Ricaldi Pit Frank. (2018). *Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión*. Universidad Nacional Daniel Alcides Carrión.
- Alfranca, Ó. (2009). Regulación ambiental e innovación. *Regulacion Ambiental e Innovación*, 15, 33–50.
- Aragón Tobón, O., Clavijo Zapata, S., Puerta Lopera, I., & Sánchez Duque, J. (2014). Formación académica, valores, empatía y comportamientos socialmente responsables en estudiantes universitarios. *Revista de La Educacion Superior*, xliii(169), 89–105. <https://doi.org/10.1016/j.resu.2015.01.003>
- Arias, F. (2012). *El Proyecto de Investigación Introducción a la metodología científica*.
- Ariza, S. M., & Marín, D. A. (2009). Factores intervinientes en la deserción escolar de la Facultad de Psicología, Fundación Universitaria Los Libertadores. *Tesis Psicológica*, 4, 72–85.
- Balagueró, T. (2018). Dataset. Retrieved April 15, 2019, from Deusto formación website: <https://www.deustoformacion.com/blog/programacion-diseno-web/que-son-datasets-dataframes-big-data>
- Barreiro, J., Vicente, E., & Pintos Clapés, J. (2005). *Introducción a Eviews*. Universidad de Vigo.
- Berzal, F., & Marín, N. (2016). Data Mining: conceptos y técnicas por Jiawei Han and Michael Kamber. *ResearchGate*, (June), 279–317. <https://doi.org/10.1016/B978-0-12-804647-0.00009-7>
- Bigdate, S. (2019). *Selección de métricas de evaluación correcta para modelos de Machine Learning: Parte 3 – Métrica de clasificación*.
- Breiman, L. (2012). Random forests. *Ensemble Machine Learning: Methods and Applications*, 157–175. https://doi.org/10.1007/9781441993267_5
- Breno, M., Peralta, I., Remache, W., & Mayorga, E. (2018). Índice de repitencia y sus causas en la Facultad de Ciencias Químicas de la Universidad Central del Ecuador. *Revenicyt*, 109–127.
- Brian, J., Flores, A., & Ponce, E. (2018). Random Forest para identificar los factores sociodemográficos asociados al uso de Internet en el Perú. *Revista ECIPeru*, 13(2), 54–61. <https://doi.org/10.33017/reveciperu2016.0008/>
- Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnología ESPOL- RTE*, 29, 170–183.
- Campos, D., & Vargas, M. (2009). *¿Cómo abordar las dificultades del aprendizaje?* Retrieved from www.oei.es/historico/idie/dificultad_aprendizaje_creditos.pdf
- Castaño, E., & Vásquez, J. (2004). *Deserción Estudiantil Universitaria: una aplicación de modelos de duración*. (60), 39–65.
- Chamba Jiménez, S. F. (2015). *Minería de Datos para segmentación de clientes en la*

empresa tecnológica Master PC. Universidad Nacional de Loja.

- Chisco Ruiz, J. T. (2015). *Asesor tutor : Tabla de contenidos*. Tecnológico de Monterrey EGE.
- Climent Hernández, J. A., Hoyos Reyes, L. F., & Martínez Preece, M. R. (2017). Formulación de un modelo híbrido alfa-estable para mercados con operación de alta frecuencia. *Contaduría y Administración*, 63(4), 58. <https://doi.org/10.22201/fca.24488410e.2018.1341>
- Condori Castro, A., & Sulla Torres, J. (2016). Comparación de Algoritmos De Clasificación para la Predicción de Casos de Obesidad Infantil. *ResearchGate*, (Abril).
- Contreras, K., Caballero, C., Palacio, J., & Perez, A. (2008). Factores asociados al fracaso académico en estudiantes universitarios de Barranquilla (Colombia). *Psicología Desde El Caribe*, 112, 110–135.
- Correa Peralta, M., Vinueza Martínez, J., Torres Arias, E., & Ponce Intriago, K. (2019). Sistema para las tutorías académicas en las universidades ecuatorianas. Caso Universidad Estatal de Milagro. *INNOVA Research Journal*, 2(6), 100–111. <https://doi.org/10.33890/innova.v2.n6.2017.219>
- Corzo Salazar, C. (2017). Deserción escolar.
- Cu Balán, G. (2008). El impacto de la escuela de procedencia del Nivel Medio Superior en el desempeño de los alumnos en el Nivel Universitario. *Revista Electrónica Iberoamericana Sobre Calidad Eficacia y Cambio En Educación*, 3(2), 59–99. <https://doi.org/2152>
- Dasso, A., & Funes, A. (2012). Agregación de métricas de Minería de Datos usando funciones de Lógica Continua. *Workshop de Investigadores En Ciencias de La Computacion*, 443–447.
- Dueñas Reyes, M. X. (2009). Minería de datos espaciales en búsqueda de la verdadera información. *Ingeniería y Universidad*, 13(1), 137–156.
- Ebriik. *Manual Microsoft Office*. , (2010).
- Echeburúa Odriozola, E. (2012). Factores de riesgo y factores de protección en la adicción a las nuevas tecnologías y redes sociales en jóvenes y adolescentes. *Revista Española de Drogodependencias*, 37(4), 435–447.
- Echeverry Marulanda, C. E., Trujillo López, M., & Mejía Salazar, M. H. (2017). Minería de datos en gestión del conocimiento de pymes de Colombia. *Revista Virtual Universidad Católica Del Norte*, 224–237.
- Eckert, K. B., & Suénaga, R. (2015). Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. *Formacion Universitaria*, 8(5), 3–12. <https://doi.org/10.4067/S0718-50062015000500002>
- Edel, R. (2003). Actitudes y comportamiento en las situaciones de conflicto. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, 1, 0.
- Equipo Editorial. (2018). Data Mining: 10 ventajas y desventajas en una empresa | Reporte Digital. Retrieved April 16, 2019, from <https://reportedigital.com/cloud/data-mining-dentro-empresa/>

- Escobar., H., Burbano., W., & Puris., A. (2016). Análisis inteligente de datos aplicado al proceso de nivelación en la Universidad Técnica Estatal de Quevedo. *Revista Publicando*, 3(7), 33–44.
- Fayrix. (2019). Selección de métricas para los modelos de aprendizaje automático.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining:towards a unifying framework. *Kdd-96*, 82–88.
- Fernández, C., Leidy, C., Alvarez Vallejo, V., Triviño Arbeláez, J., Ivan, Q., Restrepo, C. P., & Giraldo, R. (2018). *Aplicación de Técnicas de Minería de Datos para la Identificación de Patrones de Deserción Estudiantil como Apoyo a las Estrategias de Sara (Sistema De Acompañamiento para el Rendimiento Académico)*. 1177–1185.
- Fontalvo Cerpa, W., Castillo González, M. P., & Castillo, S. P. (2015). Análisis multivariado para determinar los factores más relevantes de deserción estudiantil presentes en el programa de Ingeniería Industrial de una Universidad del Caribe colombiano. *Prospectiva*, 13(1), 86. <https://doi.org/10.15665/rp.v13i1.363>
- Freire Bayas, D. E. (2014). La Conducta Disruptiva y su Incidencia en el Rendimiento Académico de las/os Estudiantes de los 10mos Grados de Educación General Básica Paralelos “I, II” del Colegio Universitario “Juan Montalvo”, de la Provincia de Tungurahua, Cantón Ambato (Vol. 593). Universidad Técnica de Ambato.
- Frites, C., & Miranda, R. (2014). Tutorías y Nivelación en la Universidad De Santiago: Tensiones y Desafíos en la Implementación de Iniciativas de Permanencia. *IV Conferencia Latinoamericana Sobre El Abandono En La Educación Superior*, 1–7.
- García, F. (2004). *Metodológicas para una encuesta*. México: Limusa.
- García Morate, D. (2008). *Manual de Weka*.
- García Muñoz, T. (2014). *Schnitzler Gut 2008 Long Term Influximab 614 Pts.Pdf* (p. 29). p. 29.
- García Rangel, E. G., García Rangel, A. K., & Reyes Angulo, J. A. (2014). Relación maestro alumno y sus implicaciones en el aprendizaje. *Ra Ximhai*, 10, 279–290. <https://doi.org/10.35197/rx.10.03.e1.2014.19.eg>
- García, S., Ramírez, S. G., Luengo, J., & Herrera, F. (2016). Big Data : Preprocesamiento y calidad de datos. *Novática*, 17–23.
- Gardner, L., Dussán, C., & Montoya, D. (2016). Aproximación causal al estudio de la deserción en la Universidad de Caldas. Periodo 2012-2014. *Revista Colombiana de Educación*, 319–340.
- Geometry, R., & Analysis, G. (2017). *II Congreso Internacional en Administración de Negocios Internacionales* (p. 778). p. 778.
- Gervilla García, E., Jiménez López, R., Montaña Moreno, J. J., Sesé Abad, A., Cajal Blasco, B., & Palmer Pol, A. (2016). La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes. *Adicciones*, 21(1), 65. <https://doi.org/10.20882/adicciones.253>
- Gironés Roid, J., Casas Roma, J., Miguinllón, A., & Caihuelas, T. (2017). *Minería de datos, Modelos y Algoritmos* (Primera Ed). Barcelona: UOC.

- Gómez Torres, F. A., Acevedo Villalobos, E. P., & Salamanca Velandia, S. R. (2015). Diagnóstico De Deserción Académica Universitaria De La Facultad De Medicina Veterinaria Y Zootecnia I. Percepción De Los Estudiantes Activos. *Puente*, 9(1), 75–84.
- González Fiegehen, L. E. (2006). Repitencia y Deserción en América Latina 2006-19. *ResearchGate*, 3, 2–40.
- Guardiola Jiménez, P. (2019). *Capítulo 1 «Estructura del SPSS»* (pp. 1–26). pp. 1–26. Retrieved from https://www.um.es/docencia/pguardio/documentos/spss_1.pdf
- Gutiérrez García, J. A. (2016). Comenzando con Weka : Filtrado y selección de subconjuntos de atributos basada en su relevancia descriptiva para la clase . *ResearchGate*, (January), 18.
- Haro, S., Pazmiño Maji, R., Conde, M., & Peñalvo, F. (2018). Minería de datos para descubrir tendencias en la clasificación de los trabajos de titulación. *Congreso de Ciencia y Tecnología ESPE*, 13(1), 125–128. <https://doi.org/10.24133/cctespe.v13i1.739>
- Himmel, E. (2018). Modelo de análisis de la deserción estudiantil en la educación superior. *ResearchGate*, 92–107. <https://doi.org/10.31619/caledu.n17.409>
- Holgado Apaza, L. A. (2018). *Detección de patrones de bajo rendimiento académico mediante técnicas de minería de datos de los estudiantes de la Universidad Nacional Amazónica de Madre de Dios*. Universidad Nacional del Altiplano.
- Hurtado Cruz, E. R. (2016). El método de mínimos cuadrados. *Facultad de Ciencias UMAN*, 1, 2–5.
- Jadue, G. (2002). Factores psicológicos que predisponen al bajo rendimiento, al fracaso y a la deserción escolar. *Estudios Pedagógicos (Valdivia)*, 193–204.
- Laguna, C. (2009). Correlación Y Regresión Lineal. *Instituto Aragonés De Ciencia De La Salud*, 1–18. Retrieved from <http://www.ics-aragon.com/cursos/salud-publica/2014/pdf/M2T04.pdf>
- Lanzarini, L. C., & Villa Monte, A. *Enseñanza De La Minería De Datos Y El Impacto De Las Herramientas De Software Utilizadas.* , (2017).
- Larrañaga, P., Inza, I., & Moujahid, A. (2010). Clasificadores Bayesianos. Universidad del País Vasco-Euskal Herriko Unibertsitatea.
- Lladó, M., & Mares, H. (2017). Factores Que Impactan La Deserción Escolar: Percepción De Los Estudiantes De La Escuela Preparatoria Federalizada No.1 Ing. Marte R. Gómez. *Comie*, (2015), 1–12.
- Marín Diazaraque, J. M. (2012). Introducción a las redes neuronales aplicadas. *Manual Data Mining*, 1–31.
- Marrón, J. A. (2008). Análisis de series temporales. *Univerisidad Complutense de Madrid*, 1(Análisis temporal), 111–128. <https://doi.org/10.3989/mc.2004.v54.i274.233>
- Martinez, A. (2018). Qué es Estado de ánimo - Psicología en Positivo. Retrieved June 23, 2019, from <https://www.antonimartinezpsicologo.com/estado-de-animo>
- Martínez, E. H., & Sanjurjo, R. L. (2014). *Minera de datos aplicada a la detección de Cáncer de Mama*.

- Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. *Revista Electronica de Investigacion Educativa*, 20(1), 38–47. <https://doi.org/10.24320/redie.2018.20.1.1347>
- Maya, E. (2014). *Métodos y técnicas de investigación* (Universidad Nacional Autónoma de México). <https://doi.org/10.1017/CBO9781107415324.004>
- Méndez Rodríguez, A. (2008). *Investigacion bibliografica.pdf*.
- Millan Reyes, L. C., Barrera Sánchez, L. F., & Ospina Diaz, J. M. (2015). Caracterización del bullying en estudiantes de medicina de Tunja , Boyacá. *Revista Virtual Universidad Católica Del Norte*, 102–112.
- Ministerio Coordinador de Patrimonio. (2015). Nacionalidades y Pueblos Indígenas y Políticas Interculturales en Ecuador. *Unicef*.
- Miranda, M. A., & Guzmán, J. (2017). Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos. *Formacion Universitaria*, 10(3), 61–68. <https://doi.org/10.4067/S0718-50062017000300007>
- Molina López, J. M., & García Herrero, J. (2006). *Técnicas de análisis de datos*. Universidad Carlos III de Madrid.
- Mosteiro García, J. (1997). *El género como factor condicionante de la elección de carrera : hacia una orientación para la igualdad de oportunidades entre los sexos*. 305–315.
- Navarro, R. (2008). El Rendimiento Académico : Concepto, Investigación Y Desarrollo. *REICE Revista Electrónica Iberoamericana Sobre Calidad Eficacia y Cambio En Educación*, 6(2), 59–99. <https://doi.org/2152>
- Osorio, A. M., Bolancé, C., & Castillo Caicedo, M. (2019). Deserción y graduación estudiantil universitaria: una aplicación de los modelos de supervivencia. *Revista Iberoamericana de Educación Superior*, iii, 31–57. <https://doi.org/10.22201/iissue.20072872e.2012.6.55>
- Oviedo C., E. A., Oviedo C., A. I., & Vélez S., G. L. (2015). Minería De Datos: Aportes Y Tendencias En El Servicio De Salud De Ciudades Inteligentes. *Revista Politécnica ISSN Año*, 11(20), 111–120.
- Padilla, A. (n.d.). El desarrollo del proceso de aprendizaje en el aula. Retrieved June 23, 2019, from https://www.tendencias21.net/ciclo/El-desarrollo-del-proceso-de-aprendizaje-en-el-aula_a41.html
- Padilla Gómez, A., López Rodríguez del Rey, M. M., & Rodríguez Morales, A. (2015). La formación del docente universitario. Concepciones teóricas y metodológicas. *Revista Científica de La Universidad de Cienfuegos*, 86–90.
- Padilla Muñoz, A. (2010). Discapacidad : contexto , concepto y modelos. *International Law, Revista Colombiana de Derecho Internacional*, 16, 381–414.
- Páramo, G. J., & Correa, C. A. (1999). *Deserción estudiantil universitaria. Conceptualización*. Colombia: Revista Universidad Eafit.
- Pastrana, K., Guitiérrez, A., & Mendoza, M. (2018). *Deserción escolar en el Instituto Nacional Julio Cesar Castillo Ubau del turno matutino en el primer semestre del año 2017 en el municipio de Condega Estelí*.

- Peña, J., Soto, V., & Calderón, U. (2016). La influencia de la familia en la deserción escolar. *Revista Mexicana de Investigación Educativa*, 21, 881–899.
- Pérez, A. M., Escobar, C. R., Toledo, M. R., Gutierrez, L. B., & Reyes, G. M. (2018). Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins. *Educação e Pesquisa*, 44(0), 1–23. <https://doi.org/10.1590/s1678-4634201844172094>
- Pérez Álvarez, P. R., & Aguilar López, D. (2017). Estudios sobre deserción académica y medidas orientadoras de prevención en la Universidad de La Laguna (España). *Paradigma*, 38(1), 48–71.
- Pérez, K., & Orozco, E. (2011). Embarazo en estudiantes de carreras universitarias. *Dictamen Libre*, 8(8), 68–72. Retrieved from file:///C:/Users/User/Downloads/475-Texto del artículo-1832-1-10-20160330 (1).pdf
- Portilla Chaves, M., Rojas Zapata, A. F., & Hernández Arteaga, I. (2014). Investigación cualitativa una reflexión desde la educación como hecho social. *Docencia Investigación Innovación*, 3, 86–100.
- Presser Carne, C. (2009). *Data mining*. 1–11.
- Quintero, G., & Martín-landrove, M. (2019). Análisis de la Difusión y Curtosis Aparentes en Imágenes por Resonancia Magnética.
- Ramírez, P. E., & Grandón, E. E. (2018). Predicción de la Deserción Académica en una Universidad Pública Chilena a través de la Clasificación basada en Árboles de Decisión con Parámetros Optimizados. *Formación Universitaria*, 11(3), 3–10. <https://doi.org/10.4067/s0718-50062018000300003>
- Rodríguez Baños, J. (2017). Simpatía, resentimiento y perdón: un análisis del rol del resentimiento en la tms de Adam Smith. *Universitas Philosophica*, 34(68), 197. <https://doi.org/10.11144/javeriana.uph34-68.srps>
- Rodríguez, V., & Chica, M. (2012). Clasificación de imágenes de satélite mediante software libre : Nuevas tendencias en algoritmos de Inteligencia Artificial. *XV Congreso Nacional de Tecnologías de La Información Geográfica, Madrid, AGE-CSIC, 19-21 de Septiembre de 2012*, 19–21.
- Ruiz, R., Garcia, J., & Pérez, M. (2014). Causas y consecuencias de la deserción escolar en el bachillerato: caso Universidad Autónoma de Sinaloa. *Revista de Investigación Educativa*, 10(5), 51–74.
- Ruiz Reina, J. L. (2018). *Evaluación de modelos*.
- Salas Picón, W. M. (2015). Revisión sobre la definición del bullying. *Poiésis*, (30), 44–50. Retrieved from <http://repository.ucc.edu.co/bitstream/ucc/965/1/1858-7309-2-PB.pdf>
- Salcedo, A., Ramírez, T., & Díaz, R. (2017). ¿Abandono o Deserción estudiantil? Una necesaria discusión conceptual. *Investigación y Postgrado*, 32(1), 63–74.
- Saldaña Villa, M., & Barriga, O. A. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile. *Revista de Ciencias Sociales(RCS)*, 16(4), 616–628. Retrieved from <https://www.redalyc.org/html/280/28016613005/>
- Santos Sharpe, A., & Carli, S. (2016). Estudios globales y locales sobre el abandono de los

- estudios universitarios. Teorías, perspectivas y nuevos abordajes. *RAES Revista Argentina de Educación Superior*, 13, 6–31. Retrieved from http://www.revistaraes.net/revistas/raes13_art1.pdf
- Sarasa, R. B., Suárez, A. R., & Sánchez, R. A. (2008). *Desarrollo de un proceso de KDD en el ámbito docente : Preparación de los datos*. 1–7.
- Sinchi, E. R. (2018). Acceso y deserción en las universidades. Alternativas de financiamiento. *Alteridad*, 13, 14.
- Tamayo Benalcázar, J. B. (2017). *Análisis Comparativo De Metodologías De Minería De Datos Y Su Aplicabilidad a La Industria De Servicios*. UDLA.
- Tamez Aguirre, P. de J. (2012). *Adicción a la Red Social de Facebook y su incidencia en el Rendimiento Académico de estudiantes de la Preparatoria 20 de la Universidad Autónoma de Nuevo León*. Universidad Autónoma de Nuevo León.
- Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A., & Alvarado-Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. *Universidad Cooperativa de Colombia*, 63–86.
- Torres, C. Z., Ramos, C. A., & Moraga, J. L. (2016). Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año , mediante minería de datos [Study of variables that influence defection in freshman university student , through data mining]. *Ciencia Amazónica: (Iquitos)*, ISSN 2221-5948, ISSN-e 2222-7431, Vol. 6, N°. 1, 2016, Págs. 73-84, 6(1), 73–84. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=5608574>
- Tuero, E., Cervero, A., Esteban, M., & Bernardo, A. (2018). ¿Por qué abandonan los alumnos universitarios? Variables de influencia en el planteamiento y consolidación del abandono. *Educacion*, 21(21(2)), 131–154.
- Unicef. (2010). *El Bullying O Acoso*. 3. Retrieved from http://www.aeped.es/sites/default/files/documentos/entrega3_bullying.pdf
- Universidad Privada TELESUP. (2019). SPSS: el software ideal para el análisis estadístico y de minería de datos | Escuela de Posgrado. Retrieved April 22, 2019, from <https://escueladeposgrado.edu.pe/blog/spss-el-software-ideal-para-el-analisis-estadistico-y-de-mineria-de-datos/>
- UTC. (n.d.). Bienestar Universitario UTC. Retrieved June 23, 2019, from <http://www.utc.edu.ec/bienestar>
- Valencia Merizalde, E. (2010). Conducta Humana y Bienestar Social. *ResearchGate*, (September 2010). <https://doi.org/10.13140/RG.2.1.4639.8485>
- Vargas, E., & Valadez, A. (2016). Calidad de la escuela, estatus económico y deserción escolar de los adolescentes mexicanos. *Revista Electrónica de Investigación Educativa*, 18(1), 82–97.
- Vásquez Hidalgo, I. (2005). Tipos de estudio y métodos de investigación. Retrieved May 1, 2019, from <https://www.gestiopolis.com/tipos-estudio-metodos-investigacion/>
- Vásquez Verdugo, J. (2016). *Modelo predictivo para estimar la deserción de estudiantes en una Institución de Educación Superior* (Universidad de Chile). Retrieved from http://repositorio.uchile.cl/bitstream/handle/2250/144169/Vásquez_Verdugo

Jonathan.pdf?sequence=1

- Vidal González, M. (2014). *El uso del Perceptrón Multicapa para la clasificación de patrones en conductas adictivas*. Universitat de les Illes Balears.
- Vieira, M. (2017). ¿Cómo ingresar a la Universidad en Ecuador? - Adipiscor.
- Villalobos, C., Alvarez, I., & Vaquera Elizabeth. (2017). *Amistades Co-Étnicas e Inter-Éticas en la adolescencia: Diferencias en la calidad, conflicto y resolución de problemas*. 113–135. <https://doi.org/10.5944/educXX1.1>
- Villanueva Morales, J. R., Lugo Rodríguez, J. J., Ramírez Buenrostro, D. O., & Ramírez Pérez, N. V. (2015). Aplicación de algoritmos de clasificación para el análisis de tejido mamario y detección de cáncer de mama. *Pistas Educativas*, (114), 260–271.
- Viteri Castro, D. A., & Uquillas Narváez, M. A. (2011). *Estudio sobre la deserción estudiantil en la Pontificia Universidad Católica del Ecuador*. Retrieved from <http://repositorio.puce.edu.ec/bitstream/handle/22000/3454/T-PUCE-3398.pdf?sequence=1>
- Vizcaino Garzón, P. A. (2008). *Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA (Waikato Environment for Knowledge Analysis)*. Universidad Nacional de Colombia.
- Weka. (2019). SMO.
- Yepes Delgado, F. L., Salazar Béltran, M., & Ceferino Martínez, E. Y. (2017). *Deserción Estudiantil en el Pregrado de la Facultad de Odontología de la Universidad de Antioquia 2004-20141*. 29, 76–95.
- Zaballa Pardo, M. (2016). *Técnicas de machine learning para la detección de la negación en textos clínicos en español*. Universidad Politécnica de Madrid.
- Zavala-Guirado, M. A., Álvarez, M., Vázquez, M., Gonzáles, I., & Bazan Ramírez, A. (2018). Factores internos, externos y bilaterales asociados con la deserción en estudiantes universitarios. *Interacciones*, 4(1), 59–69. <https://doi.org/10.24016/2018.v4n1.103>

15 ANEXOS

15.1 Anexo 1: Hoja de vida del grupo de trabajo

DATOS PERSONALES

Nombres y Apellidos: Mayra Susana Albán Taipe

Documento de identidad: 050231198-8

Fecha de Nacimiento: 6 julio de 1977

Edad: 43 años

Estado Civil: Soltera

Ciudad de Domicilio: Latacunga

Dirección: Av. José María Velazco Ibarra y pasaje Carlos Aroche Minotola

Teléfono celular: 0987773341

Email institucional: mayra.alban@utc.edu.ec

Tipo de discapacidad: Ninguno



ESTUDIOS REALIZADOS Y TÍTULOS OBTENIDOS

NIVEL	TITULO OBTENIDO	FECHA DE REGISTRO	CÓDIGO DEL REGISTRO CONESUP O SENESCYT
TERCER	Ingeniera en Informática y Sistemas Computacionales.	2005-01-13	1020-05-554410
CUARTO	Magister en Gestión de la Producción	2009-10-07	1020-09-699645
CUARTO	Aspirante Doctoral en Ingeniería en Sistemas e Informática.		

AUTORAS

Nombres y Apellidos: Jhenny Flor Masabanda Yépez

Documento de identidad: 050354752-3

Fecha de Nacimiento: 26 de julio de 1995

Lugar de Nacimiento: Moraspungo, Ecuador

Edad: 23 años

Estado Civil: Soltera

Ciudad de Domicilio: Latacunga

Dirección: Sector Mayorista, Latacunga

Celular: 0991569241

Correo Electrónico: jhenny.masabanda3@utc.edu.ec

**ESTUDIOS PRIMARIOS:**

Escuela Fiscal Mixta “Gabriela Mistral”

Moraspungo – Ecuador

ESTUDIOS SECUNDARIOS:

Unidad Educativa “Moraspungo”

Moraspungo – Ecuador

ESTUDIOS SUPERIORES:

Universidad Técnica de Cotopaxi

Actualmente cursando Décimo semestre de Ingeniería en Informática y Sistemas Computacionales

Latacunga – Ecuador

AUTORAS

Nombres y Apellidos: Carla Jhoana Zapata Rocha

Documento de identidad: 050453900-8

Fecha de Nacimiento: 25 de octubre de 1996

Lugar de Nacimiento: Latacunga, Ecuador

Edad: 22 años

Estado Civil: Soltera

Ciudad de Domicilio: Latacunga

Dirección: Joseguango Bajo

Celular: 0987590924

Correo Electrónico: carla.zapata8@utc.edu.ec

**ESTUDIOS PRIMARIOS:**

Jardín de Infantes María Montessori

Latacunga – Ecuador

Unidad Educativa Particular “La Inmaculada”

Latacunga – Ecuador

ESTUDIOS SECUNDARIOS:

Colegio de Bachillerato “Primero de Abril”

Latacunga – Ecuador

ESTUDIOS SUPERIORES:

Universidad Técnica de Cotopaxi

Actualmente cursando Décimo semestre de Ingeniería en Informática y Sistemas Computacionales

Latacunga – Ecuador

15.2 Anexo 2: Encuesta

ENCUESTA PARA DETERMINAR FACTORES QUE INCIDEN EN LA DESERCIÓN ESTUDIANTIL UNIVERSITARIA

La presente encuesta permitirá descubrir los Factores que afectan a la Deserción Estudiantil Universitaria. Esta encuesta está dirigida a las personas que tengan experiencia en temas relacionados con la deserción en las universidades y sus respuestas serán utilizadas estrictamente con fines de investigación. La encuesta está dividida en tres secciones. La sección 1 está relacionada con la caracterización de las instituciones y de la persona que llena el cuestionario. La sección 2 corresponde a preguntas que determinan la influencia de los factores que inciden en la deserción, y la sección 3 contiene preguntas que complementan al estudio.

Sección 1: Información General

1. Cédula/Pasaporte
.....
2. Género
Masculino ()
Femenino ()
3. Edad
.....
4. Carrera en la que usted estudia
Sistemas de información ()
Industrial ()
Electromecánica ()
Electricidad ()
5. Ciclo al que pertenece
Primero ()
Segundo ()
Tercero ()
Cuarto ()
Quinto ()
Sexto ()
Séptimo ()
Octavo ()
Noveno ()
Décimo ()
6. Estado civil
Soltero/a ()
Casado/a ()
Divorciado/a ()

Unión libre ()

Viudo/a ()

7. Etnia

Mestizo(a) ()

Indígena ()

Afroecuatoriano(a) ()

Blanco(a) ()

Mulato(a) ()

Otro

8. ¿Posee algún tipo de discapacidad?

Si ()

No ()

Si su respuesta es sí conteste las siguientes preguntas.

9. ¿Qué tipo de discapacidad presenta?

.....

10. ¿Qué grado de discapacidad posee?

.....

11. ¿Qué porcentaje de discapacidad (%) posee?

.....

12. ¿Qué número carné de CONADIS tiene?

.....

13. País de Origen

.....

14. Lugar de procedencia

.....

15. Lugar de residencia actual

.....

16. Estatus social

Baja ()

Media ()

Alta ()

17. Vive cerca de su universidad

Si ()

No ()

18. Tipo de Hogar

- Nuclear Funcional (Padre, madre y hermanos) ()
- Inmediato Funcional (Otros Familiares) ()
- Materno Funcional (Madre y hermanos) ()
- Nuclear Funcional (Cónyuge e hijos) ()

19. Número de miembros de familia, incluido usted

.....

20. Tipo de vivienda en que reside

- Propia ()
- Arrendada ()
- Hipotecada ()
- Cedida ()
- Otra

21. Ingresos familiares mensuales

.....

22. Origen de recursos de estudios

- Recursos propios (trabajar) ()
- Padres tutores ()
- Hermanos ()
- Crédito educativo ()
- Beca estudio ()
- Otros familiares ()

23. Si seleccionó recursos propios, conteste la siguiente pregunta. ¿Cuántas horas al día le dedica al trabajo?

- Entre 1 y 4 horas ()
- Entre 4 y 8 horas ()
- Más de 8 horas ()

24. ¿Cuál es el nivel formación del padre?

- Centro alfabetización ()
- Educación básica ()
- Educación media ()
- Superior universitaria completa ()
- Superior universitaria incompleta ()
- Posgrado maestría ()

Ninguno ()

25. ¿Cuál es el nivel de formación de la madre?

Centro alfabetización ()

Educación básica ()

Educación media ()

Superior universitaria completa ()

Superior universitaria incompleta ()

Posgrado maestría ()

Ninguno ()

26. ¿Seleccione el tipo de colegio del que procede?

Fiscal ()

Municipal ()

Particular ()

Fiscomisional ()

Otro.....
....

Sección 2: Factores de Deserción

En esta sección se presentan preguntas relacionadas a los factores que pueden influir en la deserción de los estudiantes en las universidades.

Conteste las preguntas con sí o no, de acuerdo al grado de aceptación.

27. ¿Usted eligió la carrera que estudia en base a?

Elección vocacional ()

El sistema de ingreso a la universidad le asignó esta carrera ()

Porque no tenía otra opción ()

28. ¿Considera usted, que su proceso de formación académica está preparándolo para insertarse en el campo laboral después de graduarse?

Sí ()

No ()

29. ¿Completar el proceso de nivelación recibido en la universidad le ayudó en la formación académica de primer año?

Sí ()

No ()

30. ¿La carrera que usted cursa en este momento llena sus expectativas al ser la carrera que usted deseaba estudiar?

Sí ()

No ()

31. ¿La universidad en la que estudia actualmente es la institución en la que usted quería estudiar?

Sí ()

No ()

32. ¿Considera usted, que el comportamiento negativo del docente en el aula de clases influye en la decisión de los estudiantes de abandonar la universidad?

Sí ()

No ()

33. ¿Ha sentido usted, alguna vez deseos de conducta inapropiada o respuestas emocionales inapropiadas para con sus docentes?

Si ()

No ()

34. Si su respuesta es sí, indique las razones.....
.....

35. ¿Considera usted, que el no contar con amistades dentro del aula influye en la decisión del estudiante de desertar de la universidad?

Sí ()

No ()

36. ¿Considera usted, que los espacios de bienestar estudiantil son los adecuados para cubrir las necesidades de los estudiantes?

Si ()

No ()

37. ¿Ha sido usted víctima de discriminación o bullying por parte de sus compañeros?

Sí ()

No ()

38. ¿Ha sido usted víctima de discriminación o bullying por parte de sus docentes?

Sí ()

No ()

39. ¿Las expectativas profesionales que tiene planeadas a futuro son aceptadas por su familia?

Sí ()

No ()

40. ¿Ha pensado usted en algún momento en abandonar la universidad?

Sí ()

No ()

41. Si su respuesta es sí, indique las razones.....
.....

Conteste las preguntas considerando una escala del 1 al 5 según su grado de aceptación.

42. ¿Su rendimiento académico es satisfactorio?

Nada satisfactorio ()

Poco satisfactorio ()

Medianamente satisfactorio ()

Muy satisfactorio ()

Altamente satisfactorio ()

43. Si su respuesta es nada satisfactorio, indique las razones.....

44. ¿Considera usted, que sus profesores valoran su esfuerzo y trabajo en el aula de clases?

No valorado ()

Baja valoración ()

Mediana valoración ()

Alta valoración ()

Valorado totalmente ()

45. ¿Considera usted, que un inadecuado sistema de tutorías influye negativamente en la decisión de los estudiantes de abandonar las aulas universitarias?

No influye ()

Baja influencia ()

Mediana influencia ()

Alta influencia ()

Influye totalmente ()

46. ¿Considera usted, que los problemas familiares influyen negativamente en la decisión de los estudiantes de abandonar la universidad?

No influye ()

Baja influencia ()

Mediana influencia ()

Alta influencia ()

Influye totalmente ()

47. ¿Cuál es el grado de influencia que tiene una inadecuada formación académica del docente con la deserción estudiantil universitaria?

No influye ()

Baja influencia ()

Mediana influencia ()

Alta influencia ()

Influye totalmente ()

48. ¿Cuál es el grado de influencia que tiene una inadecuada orientación del aprendizaje con la deserción estudiantil universitaria?

No influye ()

Baja influencia ()

Mediana influencia ()

Alta influencia ()

Influye totalmente ()

49. ¿Considera usted, que un inadecuado desarrollo de las cualidades humanas del docente en el aula de clases puede influir en la deserción estudiantil?

No influye ()

Baja influencia ()

Mediana influencia ()

Alta influencia ()

Influye totalmente ()

50. ¿Considera usted, que un limitado conocimiento para la utilización de tecnologías especializadas de su carrera influye negativamente en los estudiantes de abandonar la universidad?

No influye ()

Baja influencia ()

Mediana influencia ()

Alta influencia ()

Influye totalmente ()

51. ¿Considera usted, que el embarazo planificado – no planificado influye en la decisión de los estudiantes de abandonar la universidad?

	No influye	Baja influencia	Mediana influencia	Alta influencia	Influye totalmente
Planificado					
No planificado					

52. ¿Cuál es el nivel de influencia que tiene el número de hijos de los estudiantes con la deserción estudiantil universitaria?

	No influye	Baja influencia	Mediana influencia	Alta influencia	Influye totalmente
Entre 1 y 2 hijos					

Entre 2 y 3 hijos					
Más de 3 hijos					

53. ¿Cuál es el nivel de influencia que tienen los grupos que componen la comunidad universitaria respecto al factor BULLYING, con la deserción estudiantil universitaria?

	No influye	Baja influencia	Mediana influencia	Alta influencia	Influye totalmente
Bullying de estudiante a estudiante					
Bullying de docente a estudiante					

54. ¿El limitado conocimiento de la asignatura por parte del estudiante influye en su necesidad de desertar?

- No influye ()
- Baja influencia ()
- Mediana influencia ()
- Alta influencia ()
- Influye totalmente ()

Sección 3: Factores Varios

55. ¿Cuánto tiempo a la semana les dedica a las siguientes actividades?

	0 horas	entre 5-10 horas	entre 10-15	entre 15-20	entre 20-25	Más de 25 horas
Consumo de alcohol						
Juegos de azar						
Drogas						
Videojuegos						
Redes sociales						
Deportes						

56. ¿Cuántos hijos tiene usted?

- 0 hijos ()
- Entre 1 y 2 hijos ()
- Entre 2 y 3 hijos ()
- Más de 3 hijos ()

57. De los siguientes ítems ¿Cuál de estos se relaciona con su estado emocional actual que imposibiliten continuar con sus estudios universitarios?

- No cuenta con el conocimiento previo para entender el contenido de ciertas materias ()

- Problemas económicos ()
 - Problemas sentimentales ()
 - Dificultades en el aprendizaje ()
 - Problemas de rendimiento académico ()
 - No se siente motivado con sus estudios ()
 - Falta de apoyo familiar ()
 - Grandes cargas de trabajo académico ()
 - Problemas de adaptación con sus compañeros ()
 - Problemas de salud ()
 - Incapacidad para construir o mantener relaciones interpersonales con sus compañeros ()
 - Ansiedad ()
 - Timidez ()
 - Decaimiento ()
 - Depresión ()
 - Soledad ()
 - Ninguna de las anteriores ()
 - Otros
- Si usted contesto otros, indique cuáles son:.....
-

58. De los siguientes factores indique ¿Cuáles se relacionan con usted en su etapa de formación académica?

- Déficit de procesamiento de información ()
 - Incapacidad para mantener la atención ()
 - Hiperactividad ()
 - Bajo autoestima ()
 - Dificultad para expresar los sentimientos ()
 - Rechazo por parte de sus compañeros por su bajo rendimiento ()
 - Rechazo por parte de sus docentes por su bajo rendimiento ()
 - Ninguna de las anteriores ()
 - Otros
- Si usted contesto otros, indique cuáles son:.....
-

59. ¿Qué estado de ánimo presenta usted en relación con el docente en el aula de clases?

- Nerviosismo ()

- Tensión ()
- Preocupación ()
- Temor ()
- Ansiedad ()
- Satisfacción ()
- Alegría ()
- Comodidad ()
- Ninguna de las anteriores ()

Otros

Si usted contesto otros, indique cuáles son:.....
.....

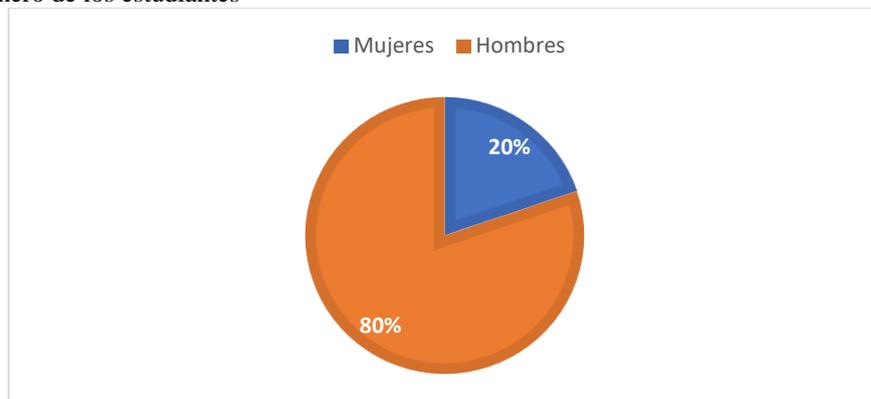
¡Este cuestionario ha finalizado, Gracias por sus respuestas!

15.3 Anexo 4: Análisis de la encuesta

Como resultado de la aplicación técnica de la encuesta, se obtuvieron los resultados como se describe a continuación:

2. Género

Figura 29: Género de los estudiantes

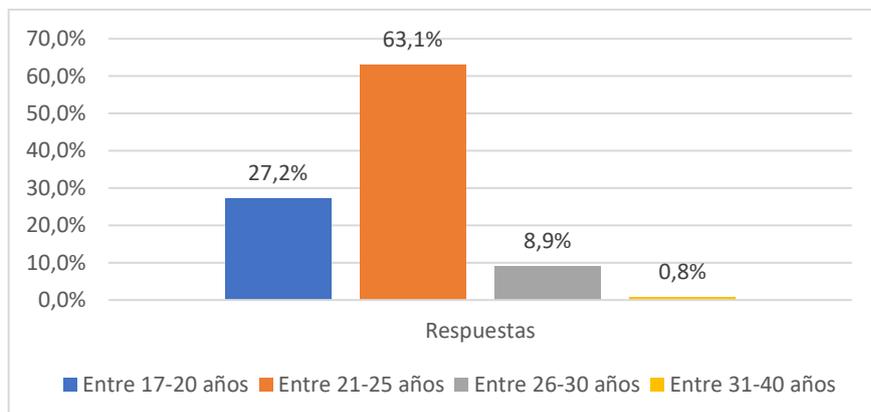


Fuente: Grupo de trabajo

En la pregunta número 2, respecto al género de los estudiantes el porcentaje con mayor índice corresponde a masculino con el 80%, mientras tanto el 20% corresponde a femenino. Según el autor (Mosteiro García, 1997), menciona que el género femenino opta por estudiar carreras profesionales identificadas con el rol tradicional de la mujer tales como: Educación Social, Enfermería, Diplomado en Relaciones Laborales, entre otras, por tal razón se considera que las mujeres optan por estudiar en Facultades y Escuelas no técnicas, mientras tanto cabe mencionar que el mayor porcentaje de estudiantes que cursan las carreras técnicas pertenece al género masculino debido a que muestran una alta autoeficacia en relación a ocupaciones tradicionales y no tradicionales.

3. Edad

Figura 30: Figura 27: Edad de los encuestados

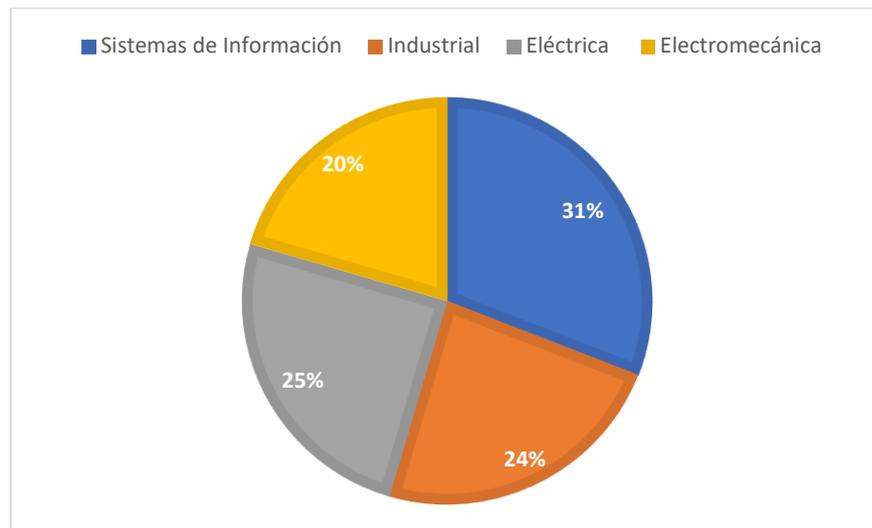


Fuente: Grupo de trabajo

Según los datos de la encuesta se puede determinar que las edades comprendidas entre (17 a 20 años) existen 397 estudiantes que corresponden al 27%, entre las edades de (21 a 25 años) 919 estudiantes que corresponde al 63%, entre las edades de (27 a 30 años) 130 estudiantes que corresponde al 9% y entre las edades de (31 a 40 años) 11 estudiantes que corresponde al 1%. Mediante la revisión de la literatura el autor (Contreras, Caballero, Palacio, & Perez, 2008), manifiesta que entre las edades comprendidas de (17 a 22 años), corresponde a la adolescencia y a la etapa del adulto joven, lo que significa que en los primeros años de universidad la población universitaria en su mayoría son adolescentes, como bien lo define la psicología evolutiva.

4. Carrera en la que usted estudia

Figura 31: Carreras de Ciencias de la Ingeniería y Aplicadas

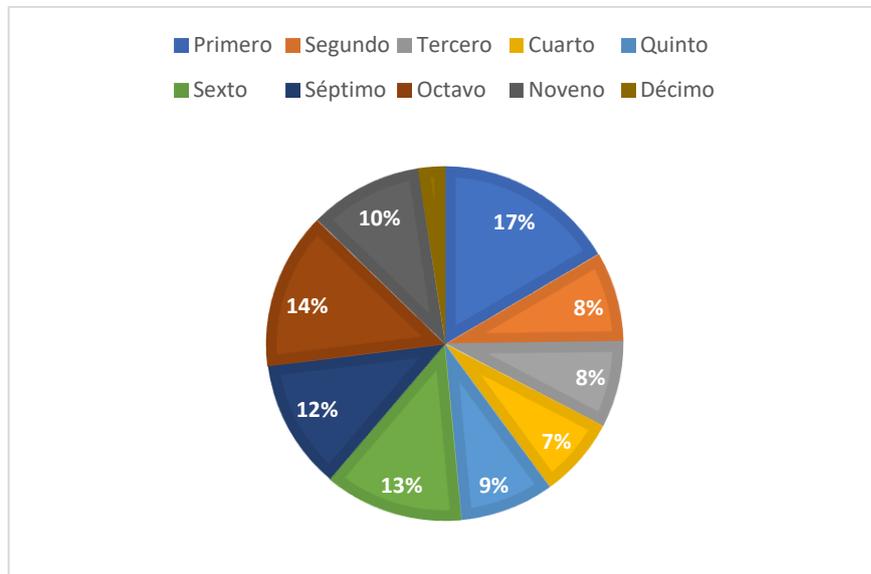


Fuente: Grupo de trabajo

Según los resultados de la encuesta se puede notar que la carrera que más ha aportado al proceso de investigación pertenece a Sistemas de información con un 31% que corresponde a 449 estudiantes, sin embargo, también se puede considerar que la carrera que menores datos ha aportado fue Electromecánica con un 20% que corresponde a 297 estudiantes.

5. Ciclo al que pertenece

Figura 32: Ciclos al que pertenece los estudiantes

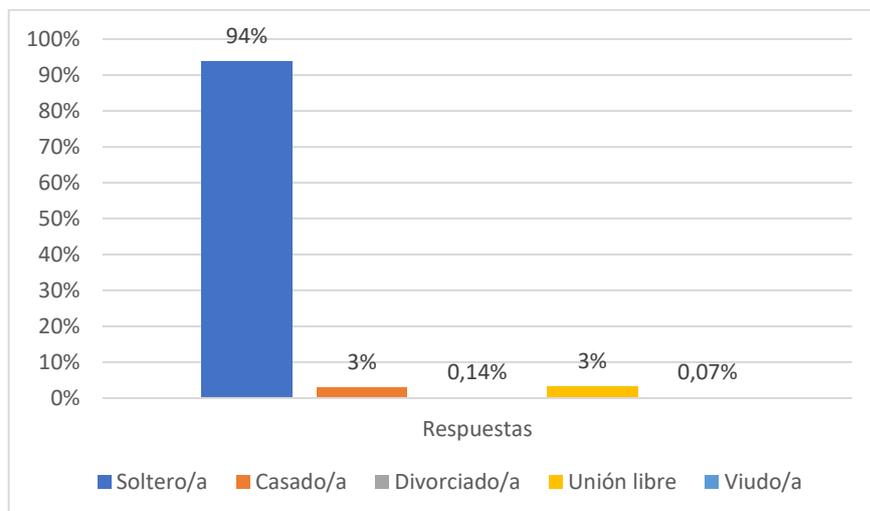


Fuente: Grupo de trabajo

En la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi, cada carrera consta de diez semestres, iniciando desde primero hasta décimo respectivamente. En los resultados se obtuvo que la mayor cantidad de población de la comunidad estudiantil universitaria que contestó la encuesta corresponde al primer ciclo con el 17% y con menor aporte los ciclos de cuarto y décimo con el 7% y 2% respectivamente

6. Estado civil

Figura 33: Estado civil de los estudiantes

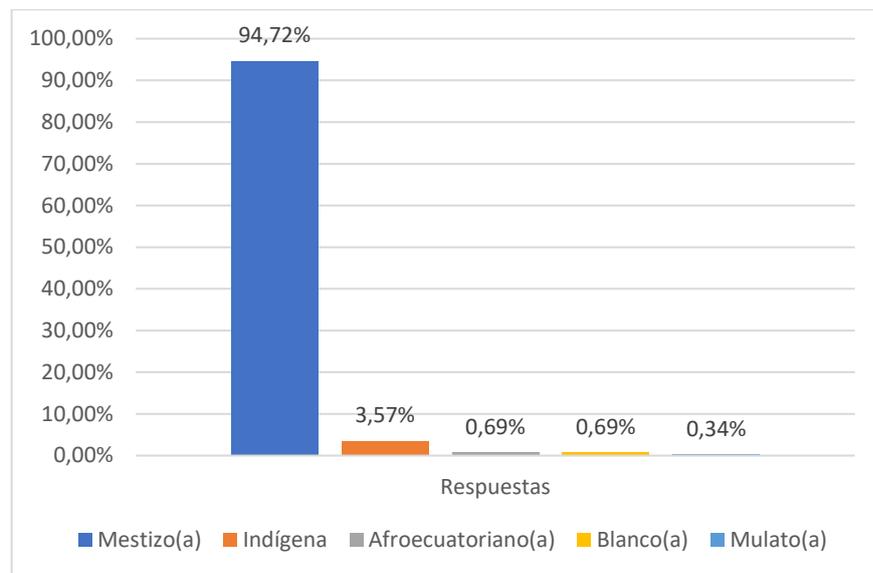


Fuente: Grupo de trabajo

De 1457 encuestados se puede determinar que el 94% de los estudiantes son solteros, el 3% corresponde a casados y unión libre y las tasas con menor porcentaje pertenece a divorciados y viudos. Cabe mencionar que los estudiantes que son casados poseen mayor riesgo de deserción, puesto que tienen más responsabilidades con su familia ya que muchas de las veces tienen que elegir entre la universidad y su familia.

7. Etnia

Figura 34: Etnia de los estudiantes

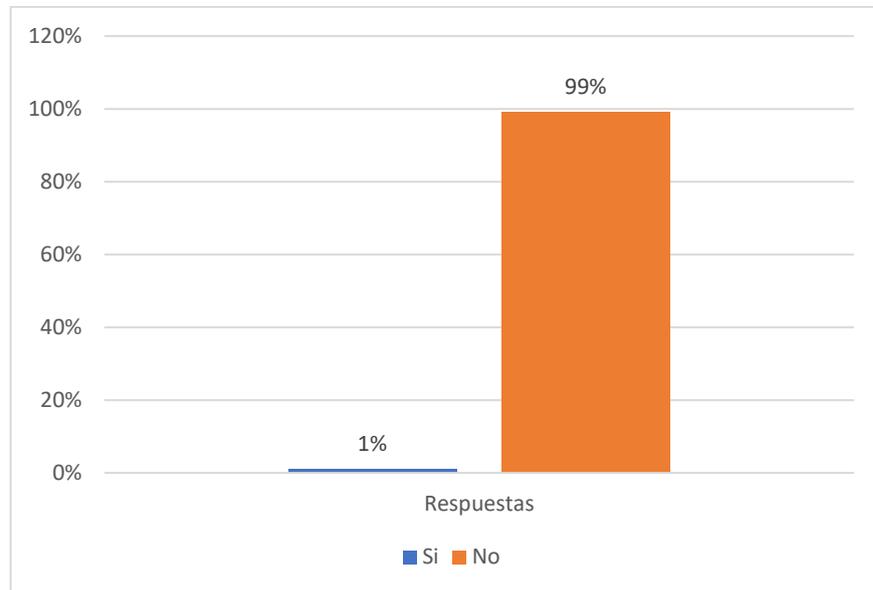


Fuente: Grupo de trabajo

Las sociedades latinoamericanas se han caracterizado por una herencia común. Desde la colonia se tipificaron a los seres humanos en castas, por ejemplo, indios, mestizos, mulatos, entre otros (Ministerio Coordinador de Patrimonio, 2015). Por tal razón según los resultados de la encuesta aplicada la mayoría de los estudiantes se consideran mestizos con el 95% que corresponde a 1380 alumnos, no obstante, se puede visualizar a 52 personas que conciernen al 4% se consideran como población indígena y el 1% como afroecuatoriano y blanco.

8. ¿Posee algún tipo de discapacidad?

Figura 35: Posee discapacidad

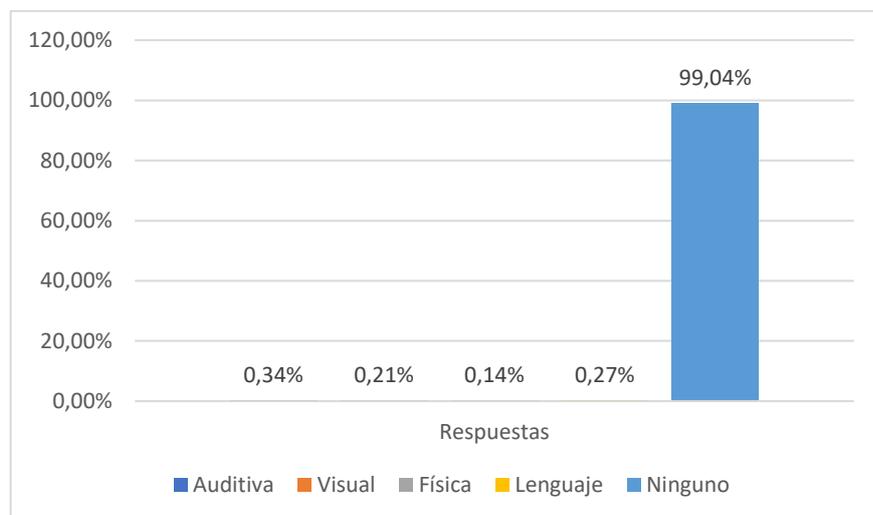


Fuente: Grupo de trabajo

La discapacidad comprende un gran número de diferentes limitaciones funcionales entre ellos, la conducta, comunicación, cuidado personal, locomoción, destrezas y aptitudes (Padilla Muñoz, 2010). Debido a la encuesta realizada se puede manifestar que de 1457 encuestados el 99% corresponde a 1443 estudiantes que no poseen discapacidad, en cambio 1% de la población encuesta es decir 14 estudiantes poseen discapacidad.

9. ¿Qué tipo de discapacidad presenta?

Figura 36: Tipo de discapacidad de posee

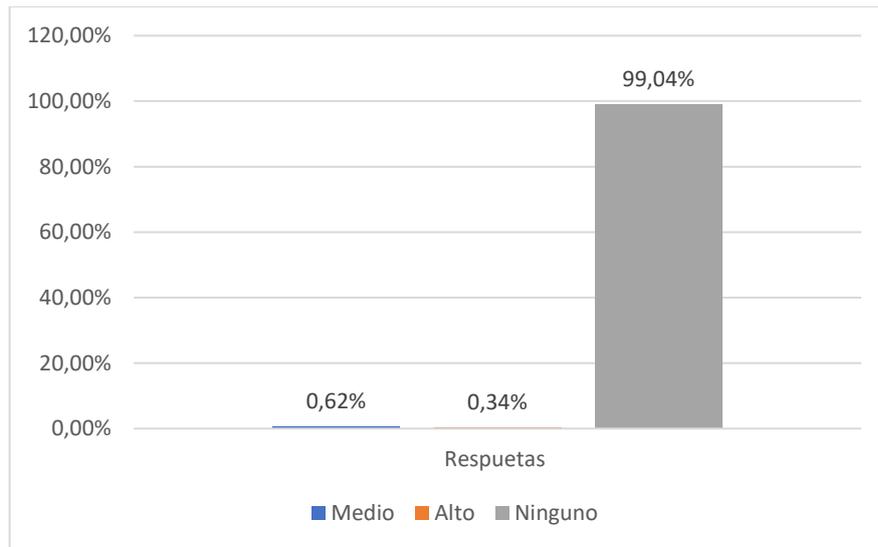


Fuente: Grupo de trabajo

Según la encuesta realizada existe un alto índice de discapacidad auditiva con el 0,34% seguida del habla o lenguaje con el 0,27%, además presenta el 0,21% de discapacidad visual y el 0,14% de discapacidad física dando como resultado un total de 14 estudiantes que poseen algún tipo de discapacidad.

10. ¿Qué grado de discapacidad posee?

Figura 37: Grado de discapacidad que posee

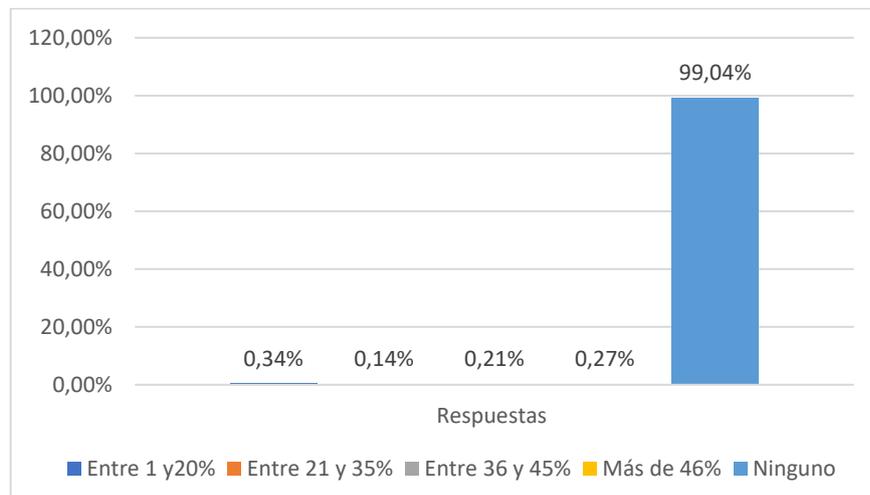


Fuente: Grupo de trabajo

Según los datos recopilados el 0,62% que pertenece a 9 estudiantes poseen un grado de discapacidad medio, el 0,34% que corresponde a 5 personas tienen un alto grado de discapacidad.

11. ¿Qué porcentaje de discapacidad (%) posee?

Figura 38: Porcentaje de discapacidad que posee

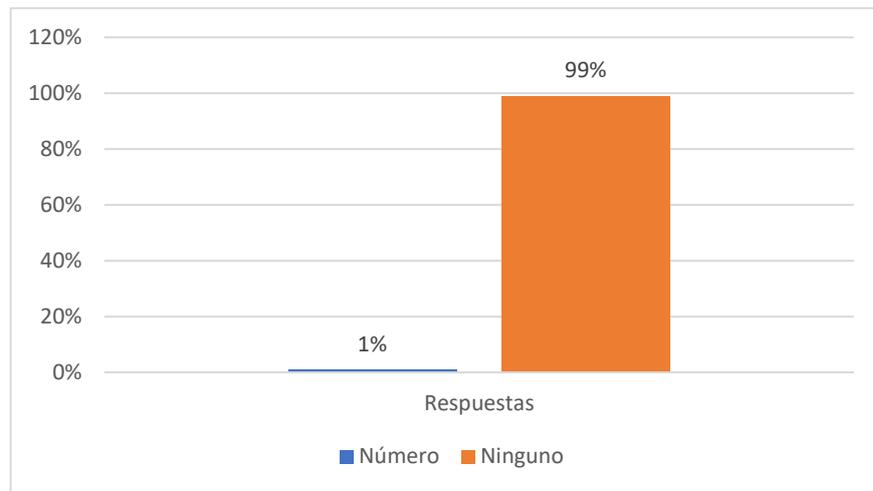


Fuente: Grupo de trabajo

Con la aplicación de la encuesta se constató que las respuestas obtenidas a la población aplicada, un 0,34% mencionaron que poseen un grado de discapacidad de (1% a 20%), el 0,14 posee un grado de discapacidad entre (21% a 35%), el 0,21 se encuentra entre (36% a 45%) y el 0,27% respondieron que tienen un grado de discapacidad superior al 46%.

12. ¿Qué número carné de CONADIS tiene?

Figura 39: Porcentaje de discapacidad que posee

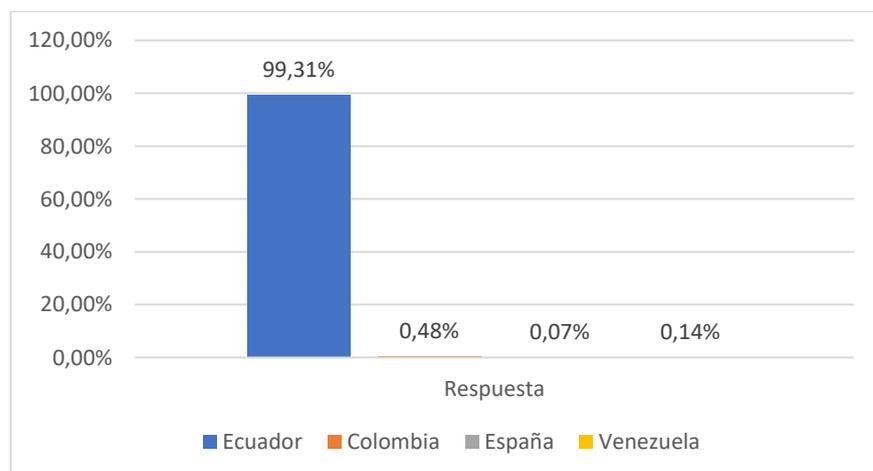


Fuente: Grupo de trabajo

Gracias a la interrogante planteada en la pregunta 12, el 1% cuentan con un carnet del CONADIS el cual les respalda en todos los trámites que realice mientras que el 99% aseguraron no contar con dicho carnet ya que no poseen ningún tipo de discapacidad.

13. País de Origen

Figura 40: País de origen de los estudiantes

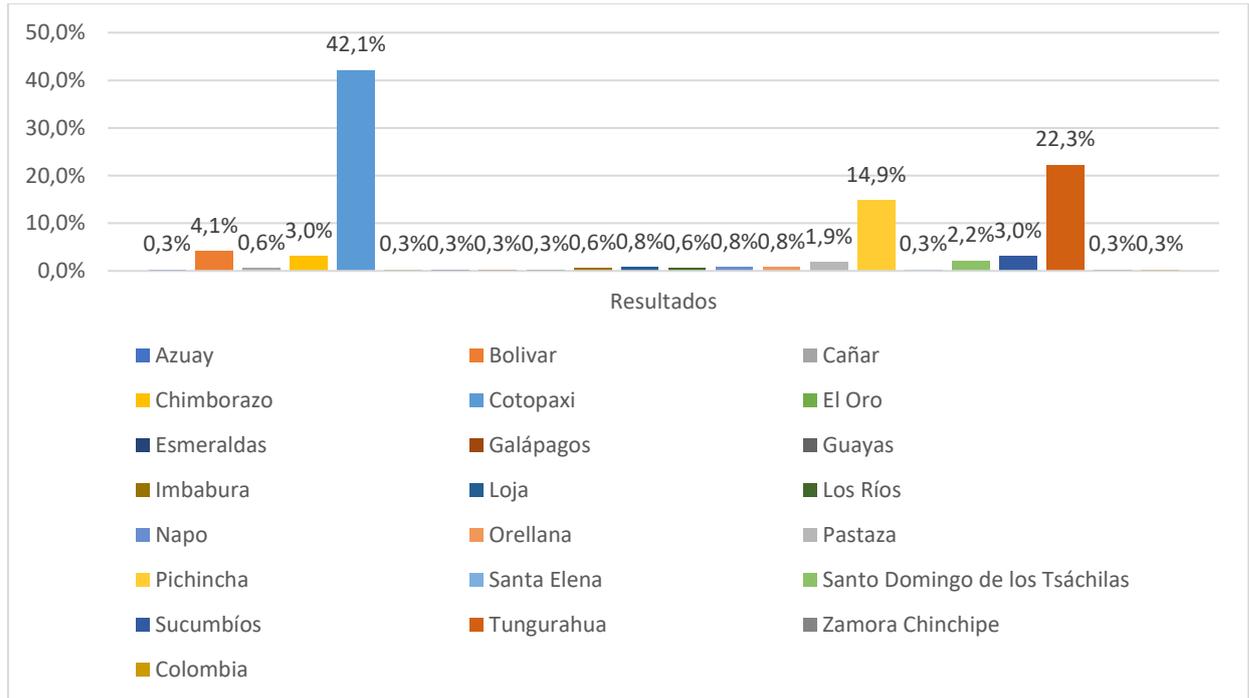


Fuente: Grupo de trabajo

Mediante los datos obtenidos de la pregunta 13, se puede evidenciar que el 99,31% de la población que corresponde a 1447 estudiantes son de nacionalidad ecuatoriana mientras que el 0,48% son de nacionalidad colombiana, el 0,07% son de nacional española y el 0,14% pertenecen a la nacionalidad venezolana.

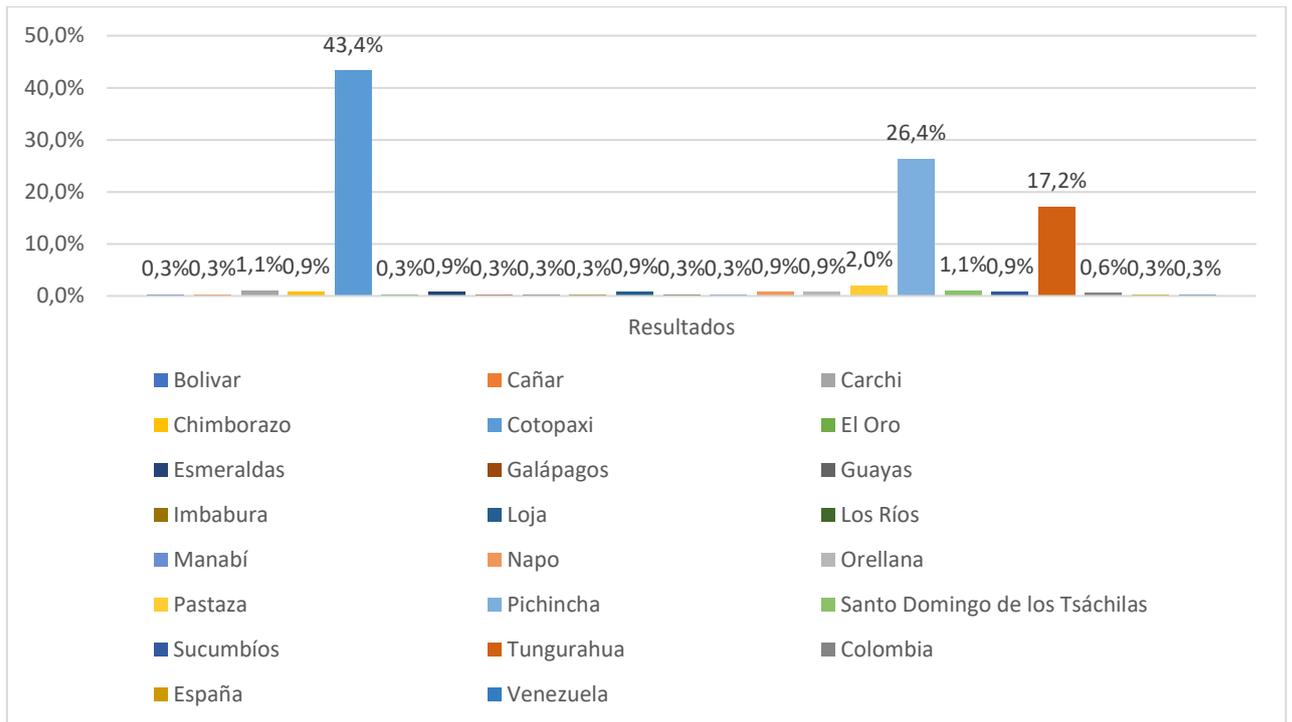
14. Lugar de procedencia

Figura 41: Lugar de procedencia de los estudiantes de Eléctrica



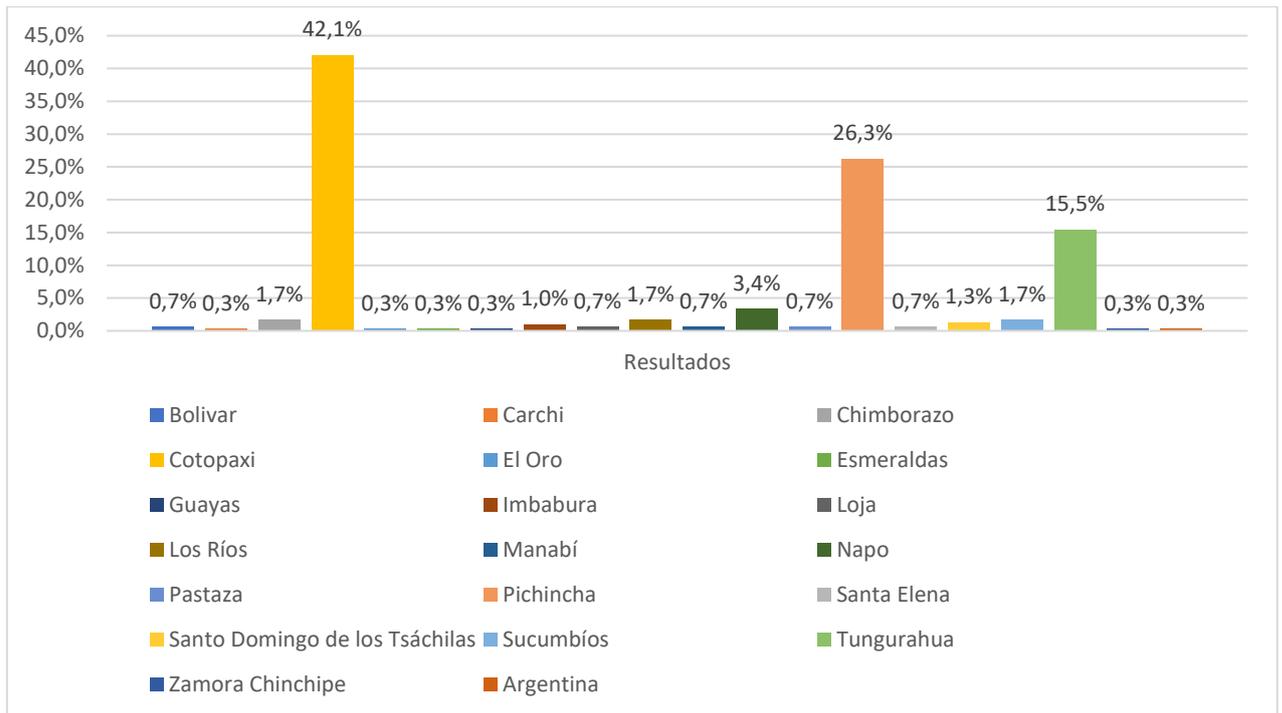
Fuente: Grupo de trabajo

Según los datos extraídos de la encuesta se realizó el análisis correspondiente sobre el lugar de procedencia de los estudiantes por carreras teniendo en primera instancia a Eléctrica en la cual el mayor índice de estudiantes son de la provincia de Cotopaxi con un 42% que corresponde a 153 estudiantes, seguida del 22% que son de la provincia de Tungurahua es decir 81 personas, con el 15% de población que pertenece a Pichincha con 54 estudiantes, no obstante existen índices bajos que provienen de las provincias de Bolívar con un 4%, Chimborazo y Sucumbíos con el 3%, Pastaza y Santo Domingo de los Tsáchilas con el 2% y con el 1% mencionan que vienen de las provincias de Cañar, Imbabura, Loja, Los Ríos, Napo y Orellana.

Figura 42: Lugar de procedencia de los estudiantes de Industrial

Fuente: Grupo de trabajo

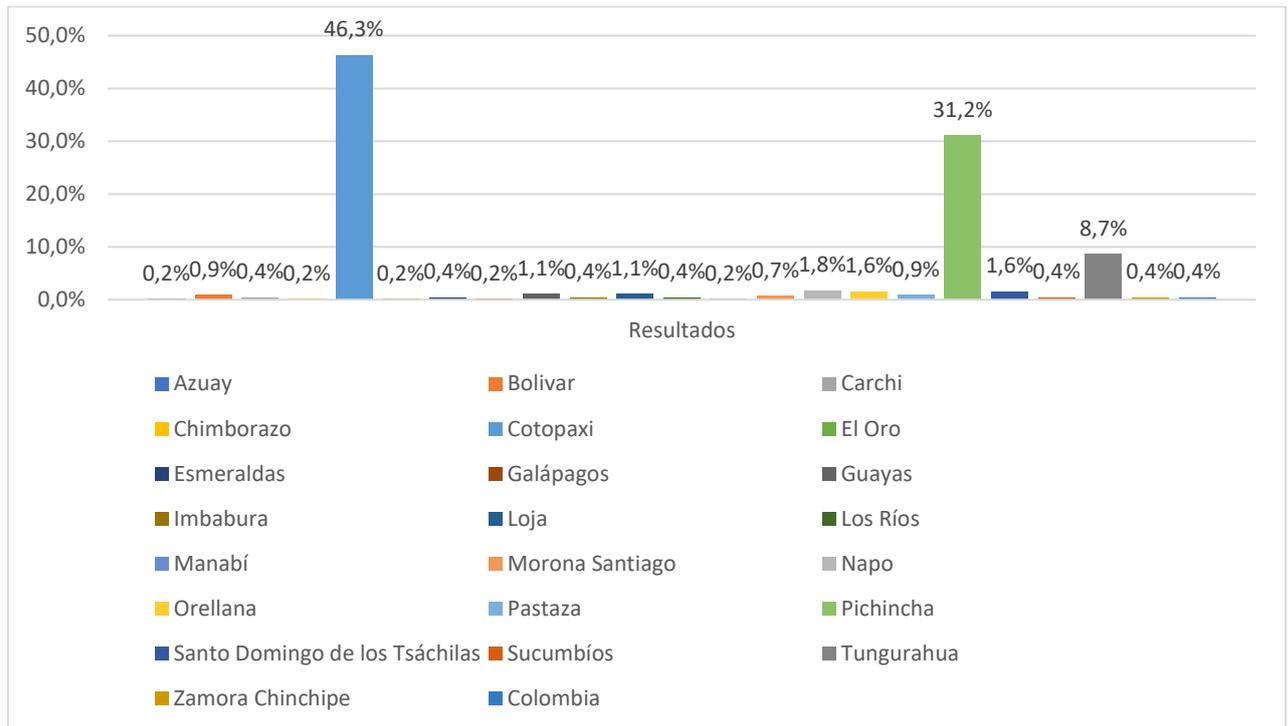
En la carrera de Industrial existe un alto índice de estudiantes que son de la provincia de Cotopaxi con 43% que pertenece a 151 personas, de Pichincha con 26% que corresponde a 92 estudiantes seguida por Tungurahua con 17% que pertenece a 60 estudiantes, además el 2% corresponde a Pastaza con 2 alumnos también presentan un índice bajo con el 1% las provincias de Carchi, Chimborazo, Esmeraldas, Loja, Napo, Orellana, Santo Domingo de los Tsáchilas, Sucumbíos y el País de Colombia.

Figura 43: Lugar de procedencia de los estudiantes de Electromecánica

Fuente: Grupo de trabajo

La carrera de Electromecánica tiene estudiantes que en su mayoría son de la provincia de Cotopaxi 125 que corresponde al 42%, en segunda instancia la provincia de Pichincha con 78 que pertenece al 26% luego la provincia de Tungurahua con 46 que hace referencia al 15%, no obstante también cuenta con estudiantes que provienen de las provincias de Chimborazo, los Ríos y Sucumbíos con el 2%, Bolívar, Imbabura, Loja, Manabí, Pastaza, Santa Elena y Santo Domingo de los Tsáchilas con el 1%.

Figura 44: Lugar de procedencia de los estudiantes de Electromecánica

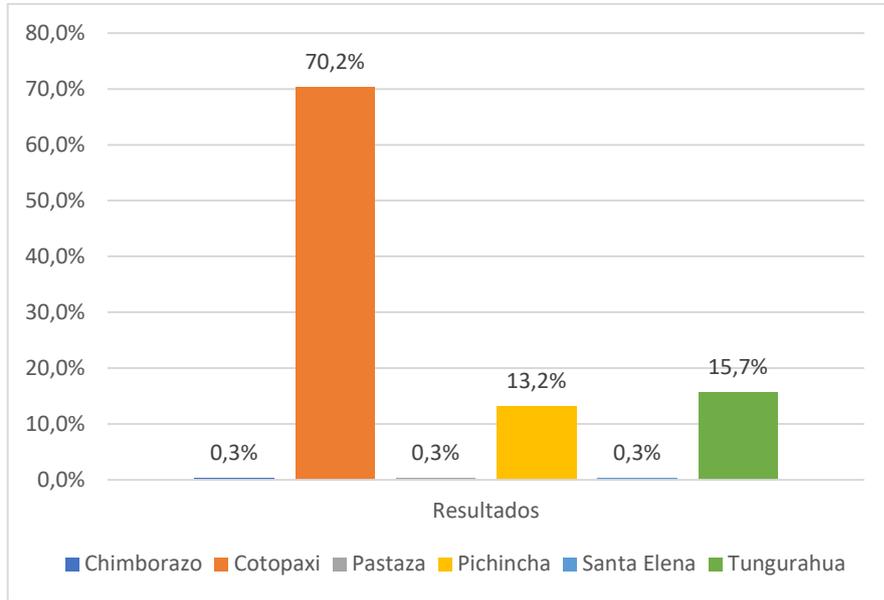


Fuente: Grupo de trabajo

Mientras tanto la carrera de Sistemas de información posee estudiantes que en su mayoría son de la provincia de Cotopaxi 208 que corresponde al 46%, en segunda instancia la provincia de Pichincha con 140 que pertenece al 31% luego la provincia de Tungurahua con 39 que hace referencia al 9%, no obstante también cuenta con estudiantes que provienen de la provincia de Pastaza con el 2%, Carchi, Chimborazo, Esmeraldas, Loja, Napo, Orellana, Santo Domingo de los Tsáchilas, Sucumbíos y el País de Colombia con el 1%.

15. Lugar de residencia actual

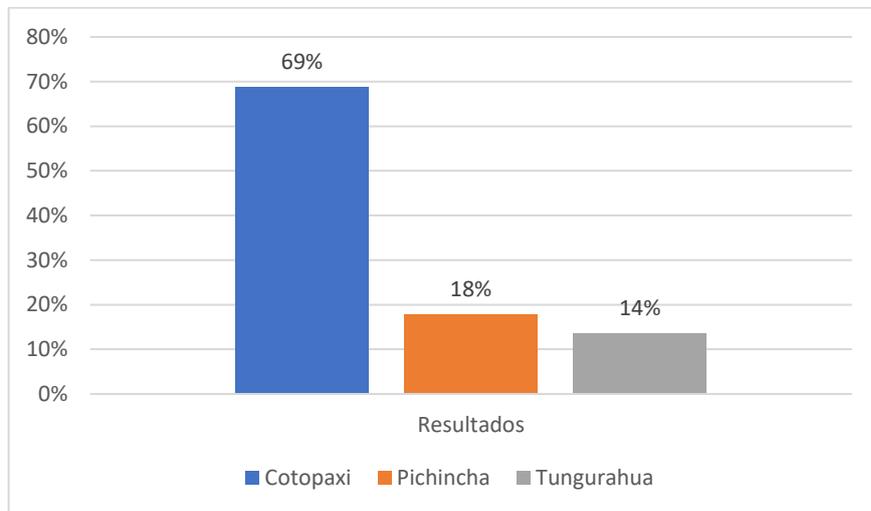
Figura 45: Lugar de residencia actual de los estudiantes de Eléctrica



Fuente: Grupo de trabajo

Según los datos extraídos de la encuesta se realizó el análisis correspondiente sobre el lugar de residencia actual de la comunidad estudiantil universitaria por carreras en la que Eléctrica cuenta con alumnos que viven en la provincia de Cotopaxi generalmente en los Cantones Latacunga, Salcedo y Pujilí con un 70% que pertenece a 255 estudiantes, seguida por la provincia de Tungurahua comúnmente de los cantones Ambato, Baños y Pillaro con un 16% que corresponde a 57, también se encuentra la provincia de Pichincha por lo regular en los cantones Quito, Mejía y Rumiñahui con 13% que se refiere a 48 estudiantes.

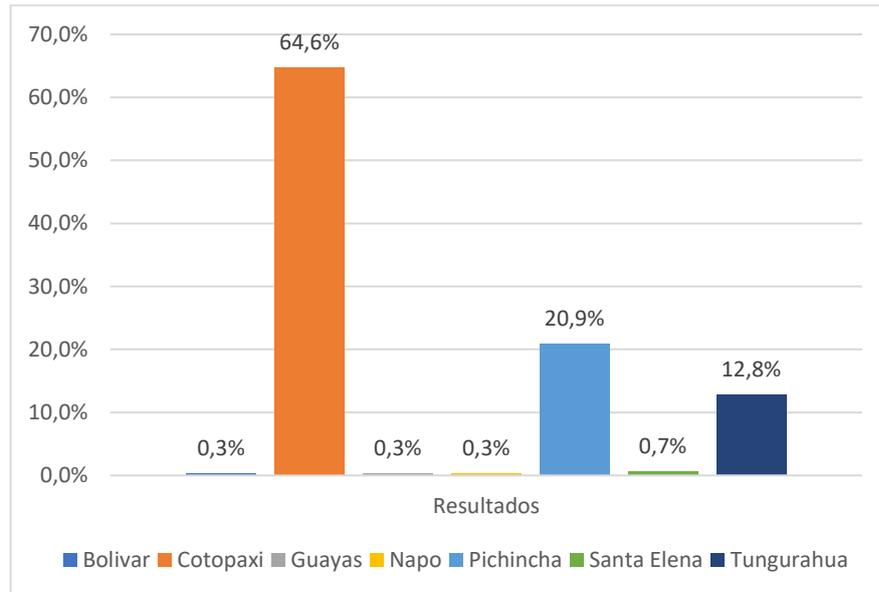
Figura 46: Lugar de residencia actual de los estudiantes de Industrial



Fuente: Grupo de trabajo

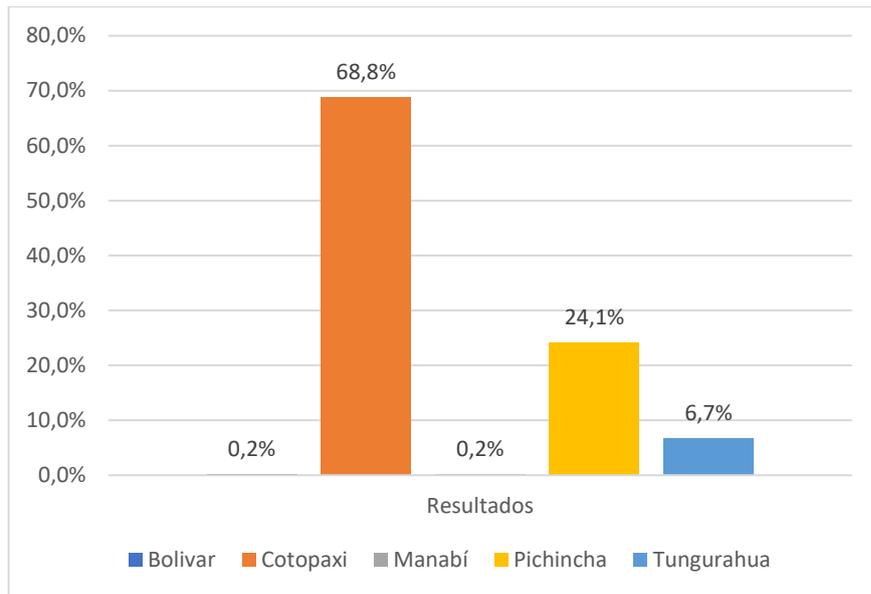
La carrera de Industrial presenta un alto índice de estudiantes que son de la provincia de Cotopaxi con 69% que pertenece a 239 personas generalmente de los cantones Latacunga y Pujilí, de Pichincha con 18% que corresponde a 63 estudiantes por lo regular de los cantones Quito, Mejía y Rumiñahui, seguida por Tungurahua con 14% que pertenece a 47 estudiantes comúnmente por los cantones Ambato, Baños y Pillaro.

Figura 47: Lugar de residencia actual de los estudiantes de Electromecánica



Fuente: Grupo de trabajo

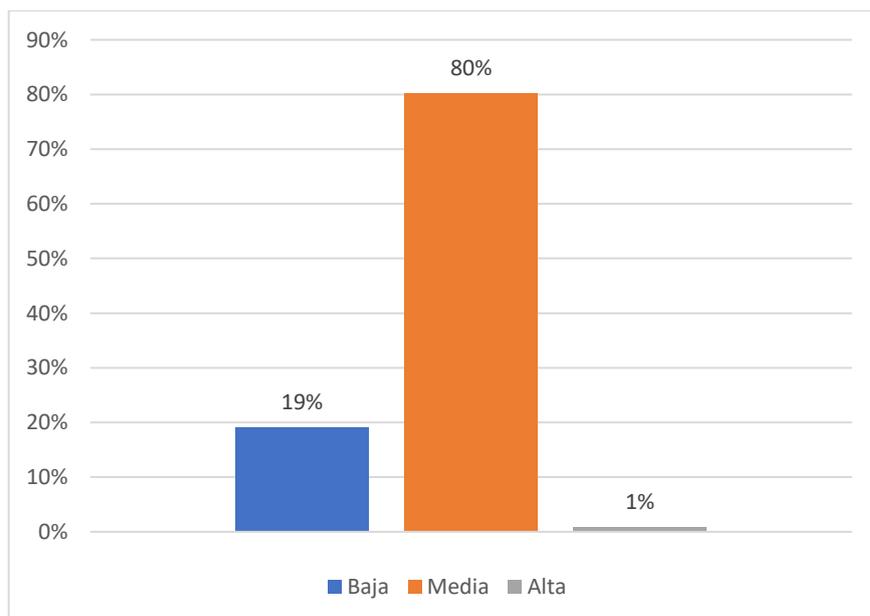
La carrera de Electromecánica tiene estudiantes que en su mayoría son de la provincia de Cotopaxi 192 que corresponde al 65% por lo general de los cantones Latacunga, Pujilí, Salcedo y Saquisilí, en segunda instancia la provincia de Pichincha con 62 que pertenece al 21% regularmente por los cantones Quito, Mejía y Rumiñahui, luego la provincia de Tungurahua con 38 que hace referencia al 13% generalmente por los cantones Ambato, Baños y Pillaro, no obstante también cuenta con estudiantes que provienen de la provincia de Santa Elena con un 1%.

Figura 48: Lugar de residencia de los estudiantes de Sistemas de Información

Fuente: Grupo de trabajo

Mientras tanto la carrera de Sistemas de información posee estudiantes que en su mayoría son de la provincia de Cotopaxi 309 que corresponde al 69% por lo general en los cantones Latacunga, Saquisilí, Pujilí, Salcedo y Sigchos, en segunda instancia la provincia de Pichincha con 108 que pertenece al 24% en los cantones Quito, Rumiñahui y Mejía, luego la provincia de Tungurahua con 30 que hace referencia al 7% en los cantones Ambato y Pillaro.

16. Estatus social

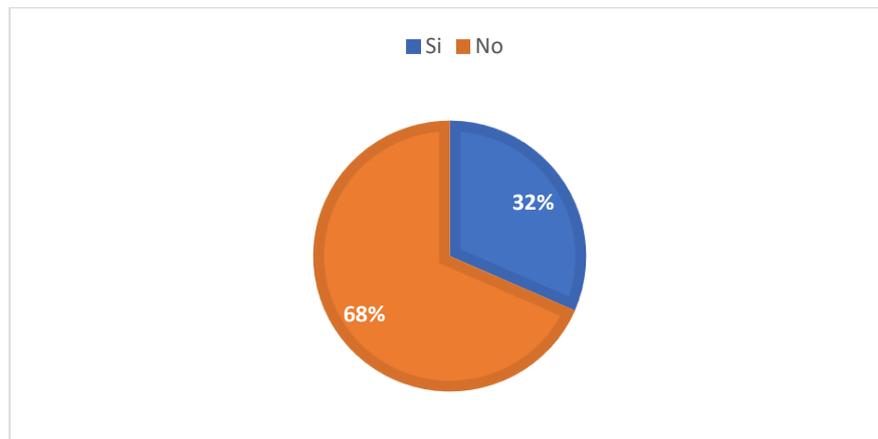
Figura 49: Estatus social de los estudiantes

Fuente: Grupo de trabajo

Una de las variables asociadas a la deserción es el factor económico que impide la continuidad del desertor en la Universidad (Páramo & Correa, 1999). Como se muestra en la interrogante 16, se puede mencionar que el 80% de la población estudiantil universitaria menciona que pertenece a un estatus medio, 19% representa a un estatus social bajo mientras tanto que el 1% de la población estudiantil universitaria se considera pertenecer a un estatus alto.

17. Vive cerca de su universidad

Figura 50: Vive cerca de la universidad

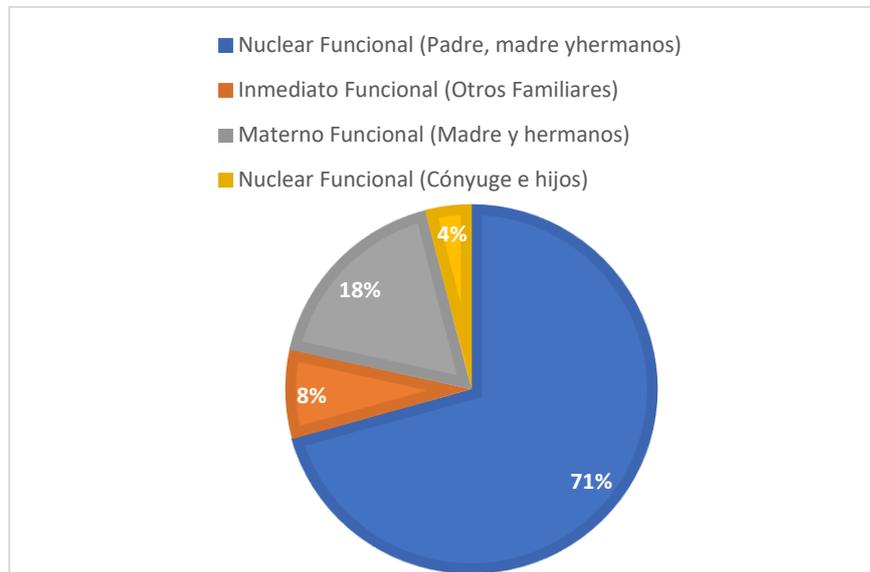


Fuente: Grupo de trabajo

En base a las respuestas obtenidas de la pregunta 17, aplicada a la comunidad académica se puede mencionar que el 68% de dicha comunidad afirman que no viven cerca de la institución educativa por lo que se le dificulta el traslado a la universidad, por otra parte, el 32% afirma que si viven cerca de la universidad.

18. Tipo de Hogar

Figura 51: Tipo de hogar de los estudiantes

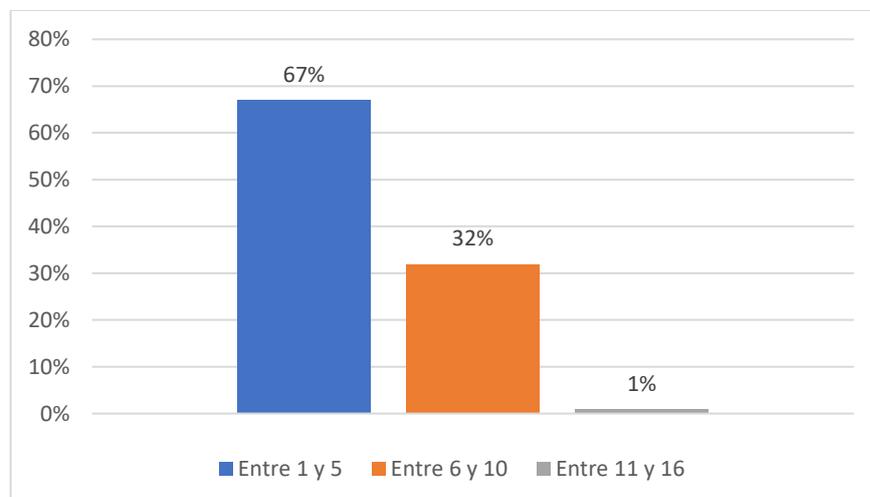


Fuente: Grupo de trabajo

Con la aplicación de la encuesta se puede afirmar que el 71% del tipo de hogar de los estudiantes está conformada por (padre, madre y hermanos), el 18% se encuentra integrada por (madre y hermanos), mientras que el 8% está constituida por (otros familiares) y el 4% por (cónyuge e hijos). Cabe mencionar que los núcleos familiares que se encuentran integradas por otros familiares y personas ajenas a la familia como por ejemplo madrastras, padrastros, entre otros, influyen de manera directa en la deserción (Yepes Delgado et al., 2017).

19. Número de miembros de familia, incluido usted

Figura 52: Número total del miembro de la familia del estudiante

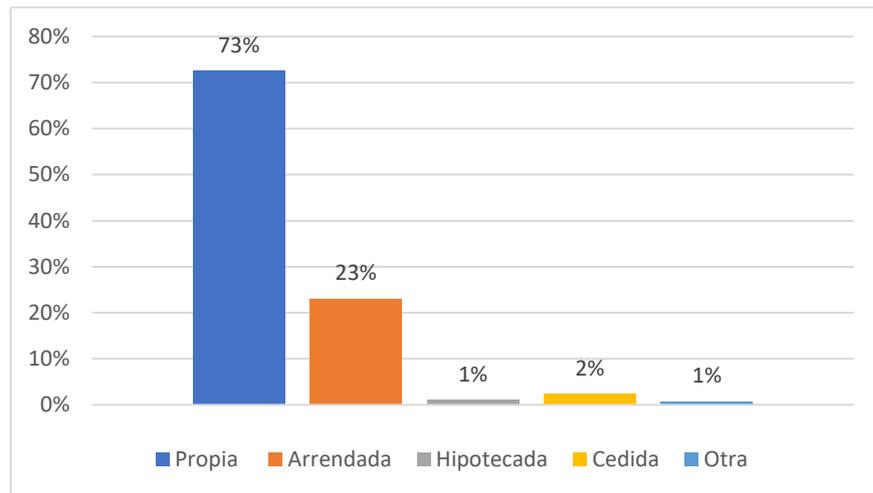


Fuente: Grupo de trabajo

Con respecto al núcleo familiar de los estudiantes se puede constatar que el 67% corresponde a (1 y 5 miembros), el 32% entre (6 y 10 miembros), mientras tanto el 1% se encuentra compuesta por (11 y 16 miembros).

20. Tipo de vivienda en que reside

Figura 53: Tipo de vivienda en la que reside

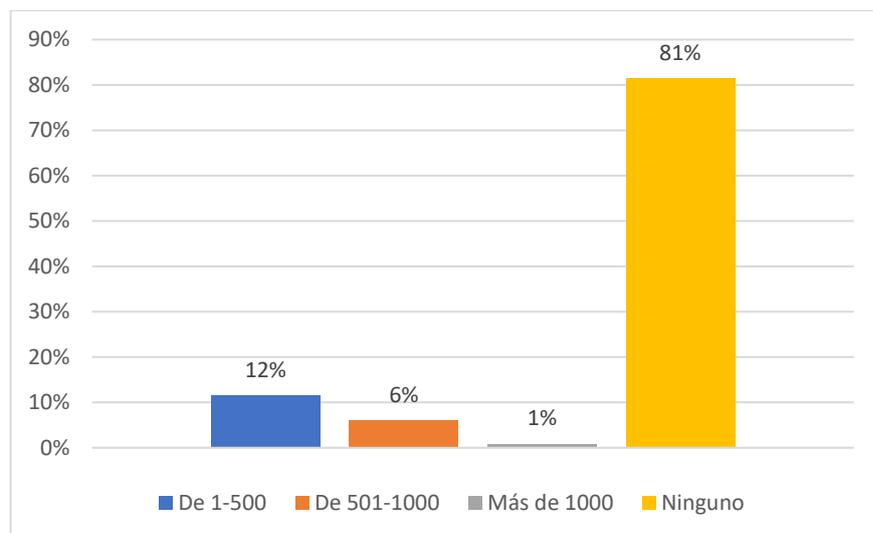


Fuente: Grupo de trabajo

En base a las resultantes de la pregunta 20, con respecto al tipo de vivienda en que reside el estudiante se obtiene que el 73% posee una casa propia, mientras tanto el 23% afirmaron que arriendan, por otro lado, el 2% mencionaron que es cedida y el 1% ente hipotecada u otra.

21. Ingresos familiares mensuales

Figura 54: Ingresos familiares mensuales

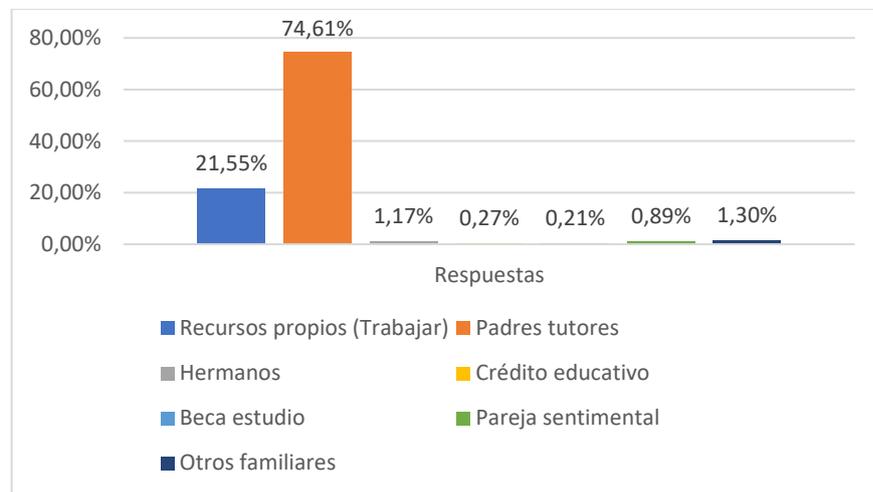


Fuente: Grupo de trabajo

Con los resultados de las encuestas se puede afirmar que el 81% de la población estudiantil no tiene un trabajo por tal razón no generan ingresos económicos, el 12% menciona que posee un ingreso entre (\$1 y \$500), mientras tanto el 6% tiene un ingreso entre (\$501 y \$1000), sin embargo el 1% de la población manifiesta que cuenta con un ingreso neto mensual superior a (\$1000) tomado en consideración que los ingresos antes mencionados corresponden a la familia del estudiante. Los ingresos familiares mensuales influyen en la deserción, debido a que el padre y la madre de familia no cuentan con ingresos económicos suficientes que permitan subsanar los gastos educativos de sus hijos o hijas.

22. Origen de recursos de estudios

Figura 55: Origen de recursos de los estudios

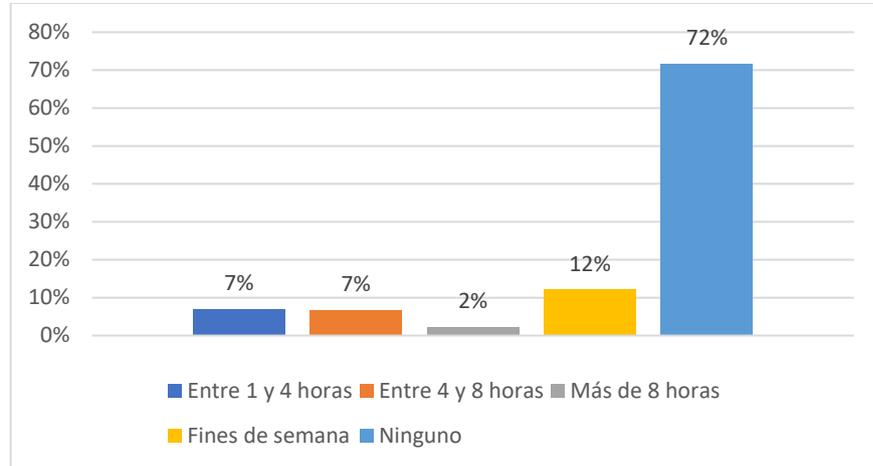


Fuente: Grupo de trabajo

Al analizar los datos generados se puede visualizar que el 74,61% del origen de los recursos de estudio depende de los padres tutores, el 21,5% son recursos propios obtenidos por el estudiante, por otro lado, el 1,17% depende de los hermanos y el 1,3% de la pareja sentimental u otros.

23. Si seleccionó recursos propios, conteste la siguiente pregunta. ¿Cuántas horas al día le dedica al trabajo?

Figura 56: Horas que se dedican a trabajar

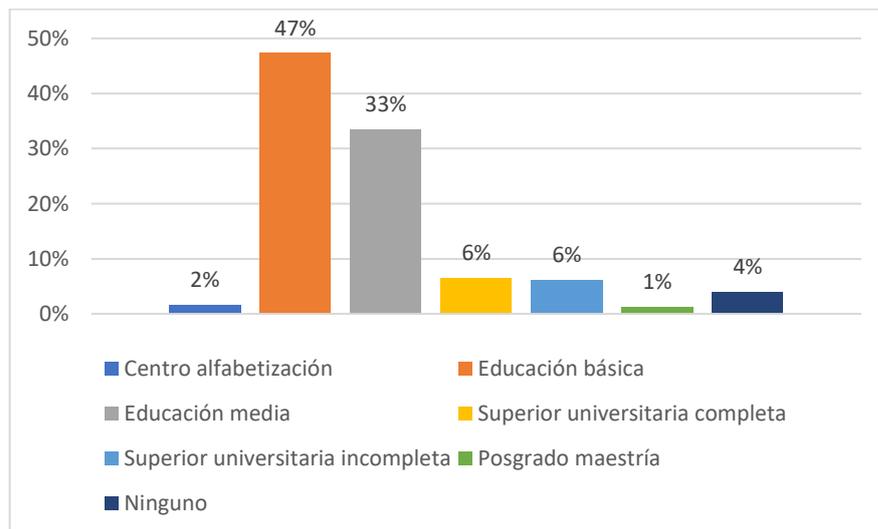


Fuente: Grupo de trabajo

Mediante los resultados obtenidos en la pregunta 23, el 12% representa a estudiantes que laboran los fines de semana, el 7% afirma que trabajan entre (1 a 4 horas) o entre (4 a 8 horas) diarias y 2% trabaja más de 8 horas.

24. ¿Cuál es el nivel formación del padre?

Figura 57: Nivel de formación del padre



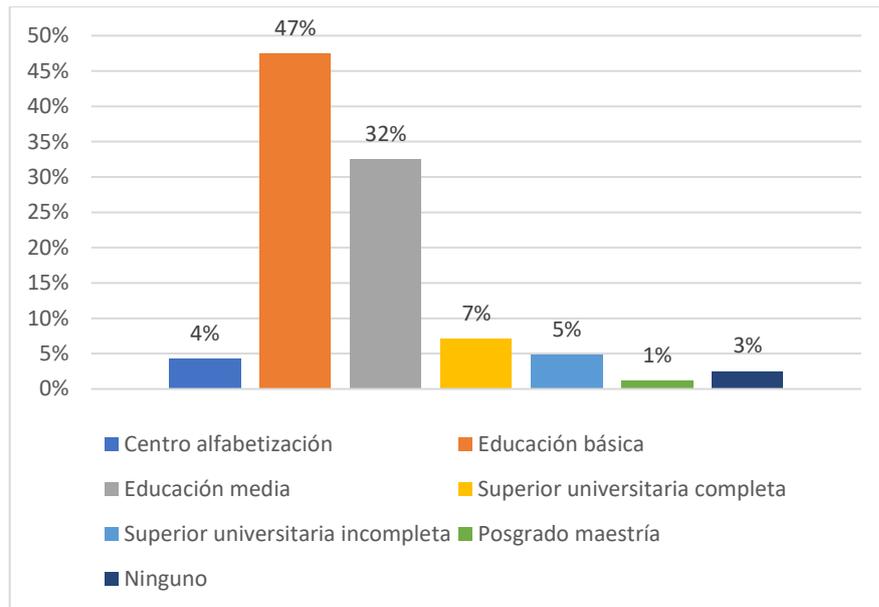
Fuente: Grupo de trabajo

En base a los resultados plasmados en la Figura de barras se puede establecer que un 47% de la población universitaria afirma que el grado de formación de sus padres es educación básica, el 33% menciona que posee educación media, el 6% mencionan que tienen una instrucción superior completa e incompleta sin embargo los bajos índices representan a los centros de

alfabetización, posgrados y ninguna. El nivel de escolaridad del padre o madre de familia de los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas puede influir en la deserción, de acuerdo al autor (Ruiz, Garcia, & Pérez, 2014), donde menciona que la deserción es común por la baja escolaridad de padres y madres.

25. ¿Cuál es el nivel de formación de la madre?

Figura 58: Nivel de formación de la madre

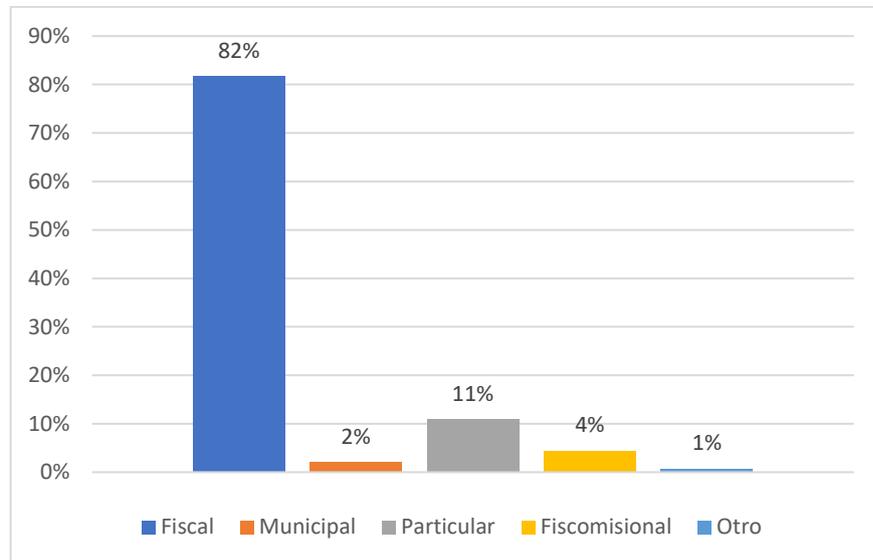


Fuente: Grupo de trabajo

El nivel de escolaridad del padre o madre de familia de los estudiantes de la Facultad de Ciencias de la Ingeniería y Aplicadas puede influir en la deserción, de acuerdo al autor (Ruiz et al., 2014), donde menciona que la deserción es común por la baja escolaridad de padres y madres. En base a los resultados plasmados en la Figura de barras se puede establecer que un 47% de la población universitaria afirma que el grado de formación de sus madres es educación básica, el 32% menciona que posee educación media, el 7% mencionan que tienen una instrucción superior completa sin embargo los bajos índices representan a los centros de alfabetización, instrucción superior incompleta, posgrados y ninguna.

26. ¿Seleccione el tipo de colegio del que procede?

Figura 59: Colegio del que proceden

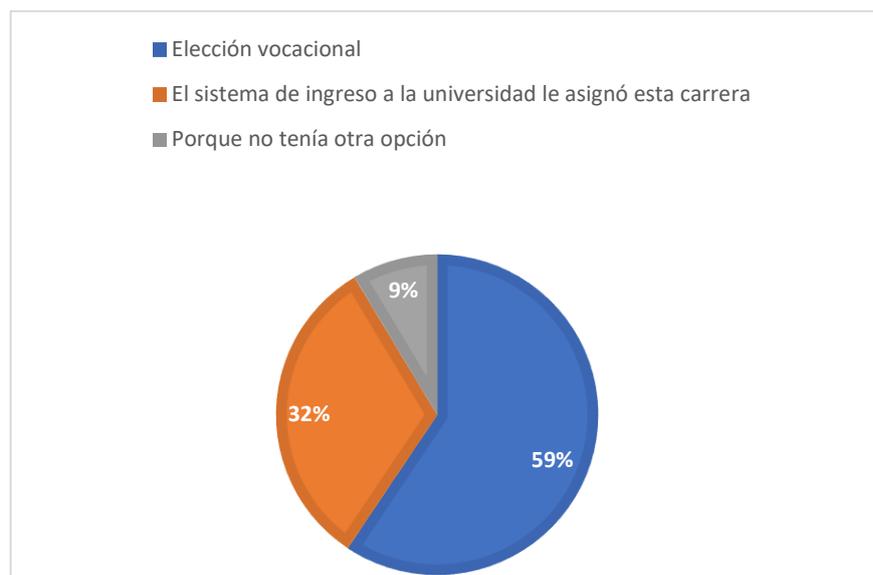


Fuente: Grupo de trabajo

Los resultados de la encuesta manifiesta que el 82% de los estudiantes afirmaron que provienen de instituciones fiscales siendo este el mayor porcentaje y el 11% menciona que culminaron sus estudios en una entidad particular, el 4% y 2% en unidades educativas fisco-misional y municipales respectivamente.

27. ¿Usted eligió la carrera que estudia en base a?

Figura 60: Elección de carrera

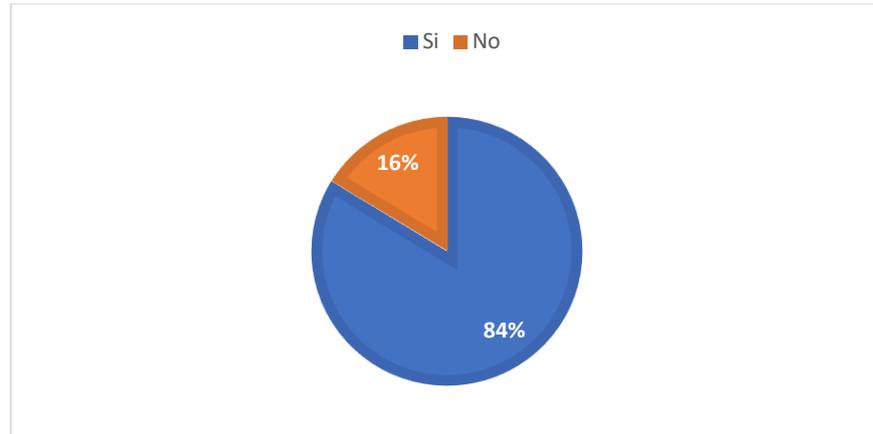


Fuente: Grupo de trabajo

Con respecto a esta interrogante el 59% de la población universitaria afirma que la carrera que está cursando la seleccionaron por elección vocacional, mientras que el 32% por el sistema de ingreso a la universidad y el 9% porque no tenían opción.

28. ¿Considera usted, que su proceso de formación académica está preparándolo para insertarse en el campo laboral después de graduarse?

Figura 61: Proceso de formación académica

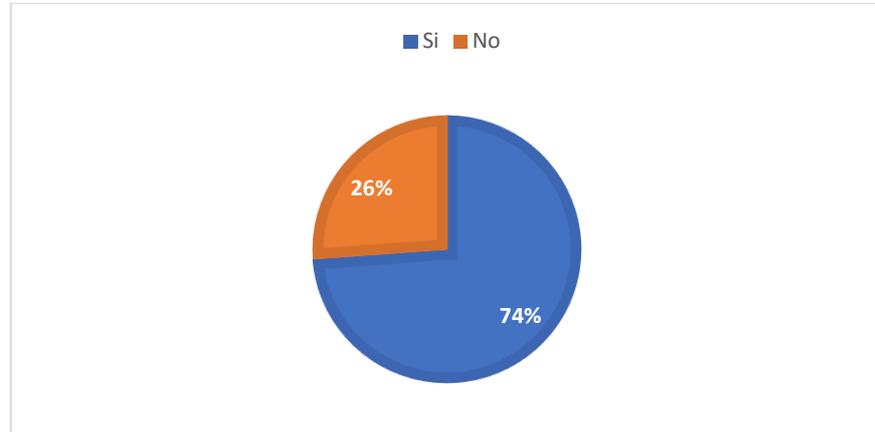


Fuente: Grupo de trabajo

La calidad escolar como factor asociado a la deserción, las profundas desigualdades socioeconómicas de los estudiantes influye directamente en la probabilidad de asistencia, continuidad y logro escolar. Por lo tanto, el riesgo de deserción en los estudiantes que asisten a una institución educativa de determinada calidad escolar varía según el nivel económico del estudiante (Vargas & Valadez, 2016). De acuerdo a la encuesta aplicada se puede mencionar que 1220 que corresponde al 84% de los estudiantes consideran que el proceso de formación académica que se encuentran recibiendo los está preparando para insertarse en el campo laboral, mientras tanto 237 que comprende el 16% manifiestan que el proceso de formación académica no es adecuado.

29. ¿Completar el proceso de nivelación recibido en la universidad le ayudó en la formación académica de primer año?

Figura 62: Proceso de nivelación

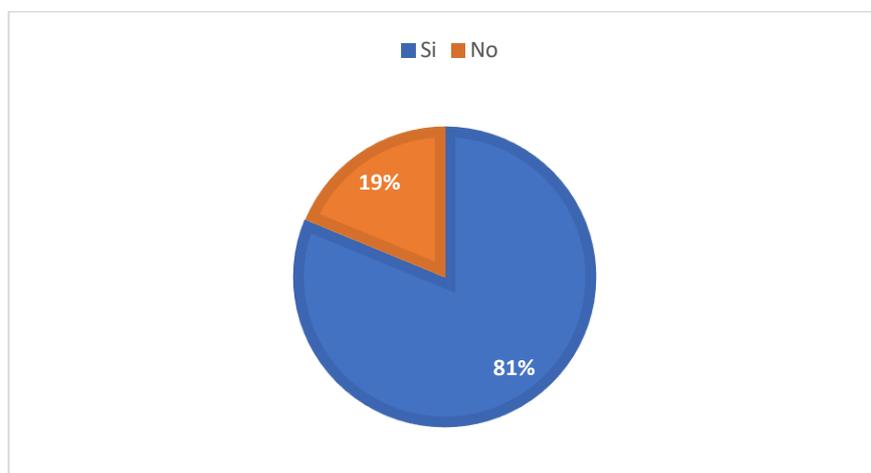


Fuente: Grupo de trabajo

Según los resultados de la encuesta el 74% de la población universitaria que comprende 1076 estudiantes manifestaron que el haber recibido nivelación le ayudó en la formación académica de primer semestre, sin embargo, el 26% que corresponde a 381 alumnos mencionaron que el proceso de nivelación recibido no contribuyó en el desarrollo académico, no obstante se debe tener en cuenta que en algunos casos los estudiantes respondieron que no aportó en la formación académica puesto que fueron exonerados de éste proceso.

30. ¿La carrera que usted cursa en este momento llena sus expectativas al ser la carrera que usted deseaba estudiar?

Figura 63: Expectativas de la carrera

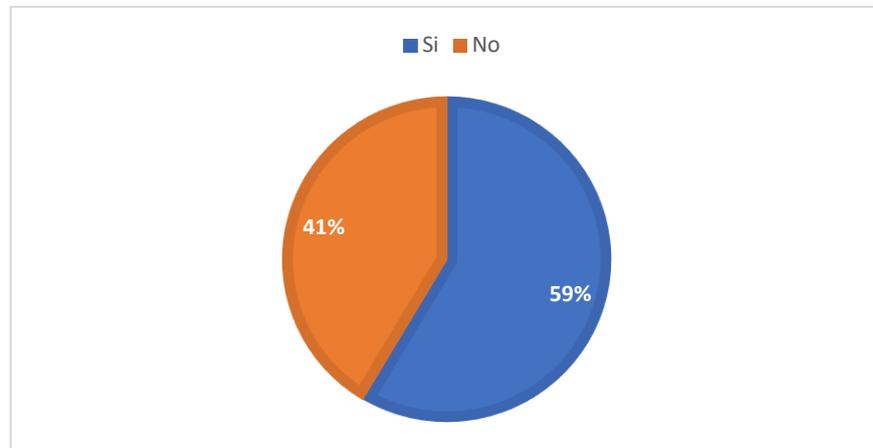


Fuente: Grupo de trabajo

Uno de los factores que puede influir en la deserción, es la mala elección de la carrera, es decir hace referencia al proceso de admisión, donde, por ejemplo, la falta de información adecuada y veraz del programa académico y de la institución al estudiante puede conducir a su deserción (Fontalvo Cerpa et al., 2015). Por otro lado, en los resultados de la encuesta se puede manifestar que el 81% de la comunidad estudiantil universitaria se siente satisfecho con la carrera que cursa actualmente, mientras tanto el 19% no se siente conforme con la misma lo cual puede ser una limitante para continuar con sus estudios.

31. ¿La universidad en la que estudia actualmente es la institución en la que usted quería estudiar?

Figura 64: Expectativas de la carrera

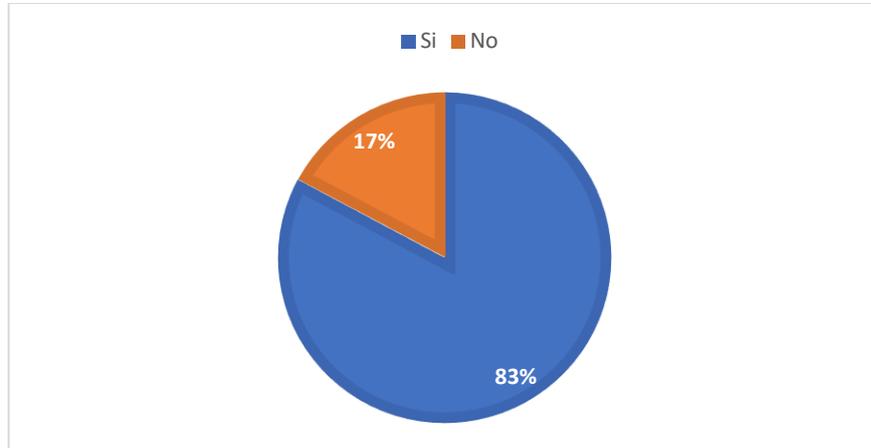


Fuente: Grupo de trabajo

De acuerdo a los datos obtenidos en la encuesta se pudo determinar que la mayoría de los estudiantes que corresponde a un 59% no querían estudiar en la universidad que actualmente se encuentran pues hubieran querido ingresar en otras universidades que existen en el país siendo su mayoría universidades públicas, pues solo los que tiene una posibilidad económica ingresan a universidades privadas, mientras que un 41% de estudiantes deseaban estudiar en la Universidad Técnica de Cotopaxi.

32. ¿Considera usted, que el comportamiento negativo del docente en el aula de clases influye en la decisión de los estudiantes de abandonar la universidad?

Figura 65: Comportamiento negativo de docentes en el aula de clases

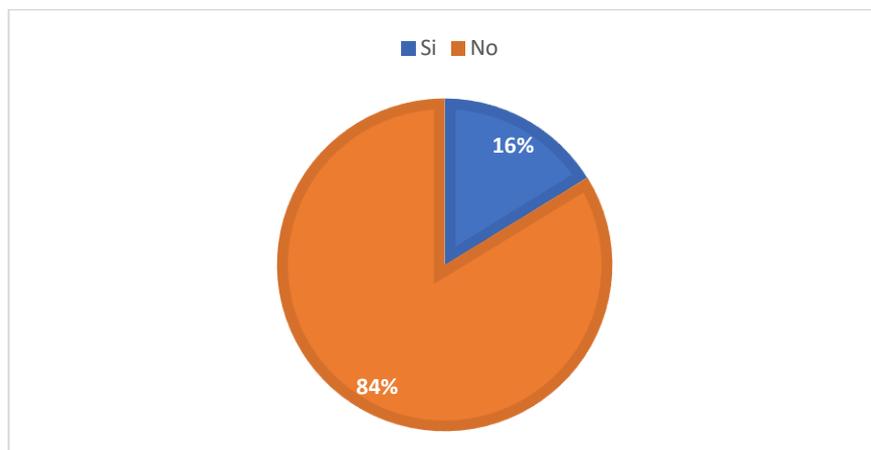


Fuente: Grupo de trabajo

El factor comportamiento está ligada a la actitud y cuando existe más presión del tiempo puede elevarse este vínculo que genere este tipo de comportamiento (Edel, 2003). Por tal razón, la mayoría de los estudiantes que son un 83% consideran que el comportamiento negativo que un docente puede tener en el aula de clases respecto con sus alumnos genera que los mismos estudiantes piensen en abandonar sus estudios universitarios y un porcentaje mínimo que corresponde al 17% piensa que no es así.

33. ¿Ha sentido usted, alguna vez deseos de conducta inapropiada o respuestas emocionales inapropiadas para con sus docentes?

Figura 66: Comportamiento negativo de docentes en el aula de clases

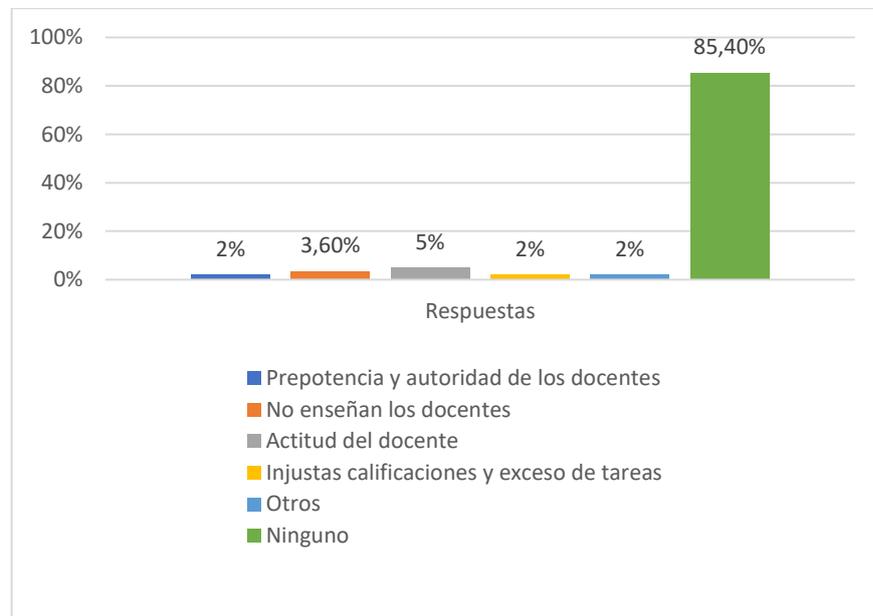


Fuente: Grupo de trabajo

Uno de los motivos para que exista conductas inapropiadas es la injusticia la misma que se da por el resentimiento o la venganza que puede existir hacia una persona, además que no soportan la conducta de aquella persona a la que tiene resentimiento (Rodríguez Baños, 2017). Por otra parte, las encuestas que se realizaron en su mayoría los estudiantes no han sentido deseos de conducta inapropiada o respuestas negativas para con los docentes que imparten sus conocimientos con los alumnos, esto corresponde a un 84% de las respuestas mientras que un 16% piensa lo contrario por diversas razones.

34.- Si su respuesta es sí, indique las razones

Figura 67: Razones de conductas inapropiadas

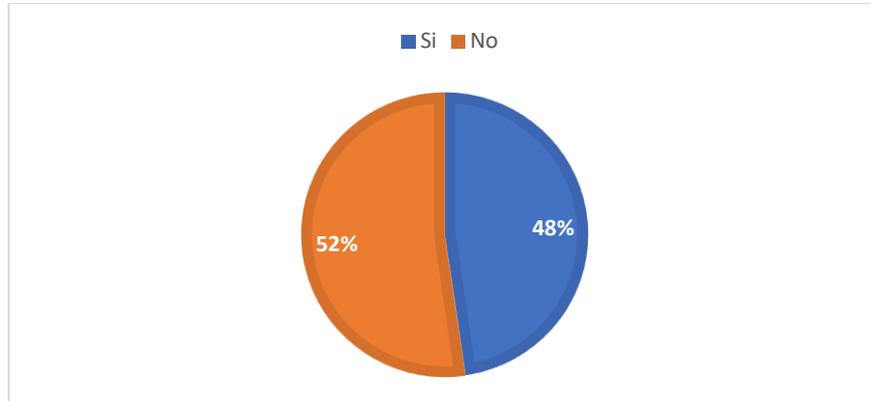


Fuente: Grupo de trabajo

En base a la pregunta 33 y de acuerdo a las respuestas, existen diversas razones y motivos que han hecho sentir a los alumnos deseos de conducta inapropiada para con los docentes entre ellas se encuentra la prepotencia, autoridad, injustas calificaciones, exceso de tareas todo esto corresponde a un 2% que es una mínima parte, mientras que un 3,6% es porque no enseñan y un mayor porcentaje de 5% que es por la actitud que presenta el docente en el aula de clases.

34. ¿Considera usted, que el no contar con amistades dentro del aula influye en la decisión del estudiante de desertar de la universidad?

Figura 68: No tener amistades

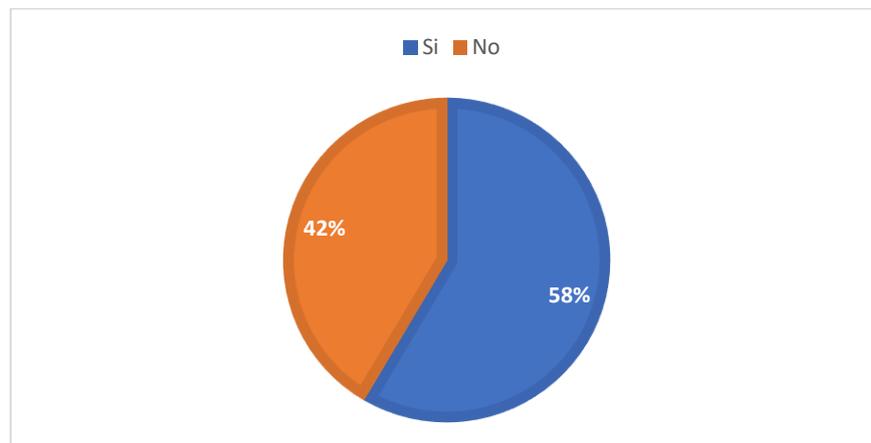


Fuente: Grupo de trabajo

Contar con una amistad te mantiene positivo, mejorar tu autoestima, bienestar psicológico y permite desarrollar la personalidad además del crecimiento socioeconómico (Villalobos, Alvarez, & Vaquera Elizabeth, 2017). Sin embargo, la mayoría de los estudiantes piensan que el no tener amistades dentro de la universidad para compartir no incide en que una persona que está cursando sus estudios académicos abandone la universidad esto corresponde a un 52%, mientras que un 48 % piensa lo contrario.

35. ¿Considera usted, que los espacios de bienestar estudiantil son los adecuados para cubrir las necesidades de los estudiantes?

Figura 69: Espacios de bienestar estudiantil adecuados

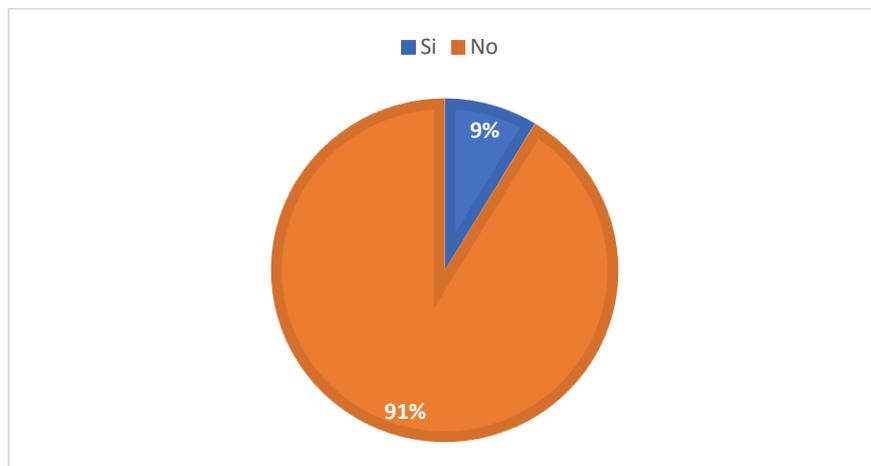


Fuente: Grupo de trabajo

Bienestar estudiantil de la Universidad Técnica de Cotopaxi tiene la misión de ofrecer servicios que sean gratuitos, el mismo que debe cuidar el bienestar bio-psico-social de los estudiantes que forman parte de la universidad (UTC, n.d.). Por tal razón, la encuesta aplicada y de acuerdo a los resultados obtenidos un 58% de los estudiantes, consideran que los espacios que brinda bienestar estudiantil a toda la comunidad universitaria están acorde a cubrir las diferentes situaciones de los estudiantes y un 42% no cree que bienestar estudiantil tenga la capacidad de resolver las necesidades de los estudiantes.

36. ¿Ha sido víctima de discriminación o bullying por parte de sus compañeros?

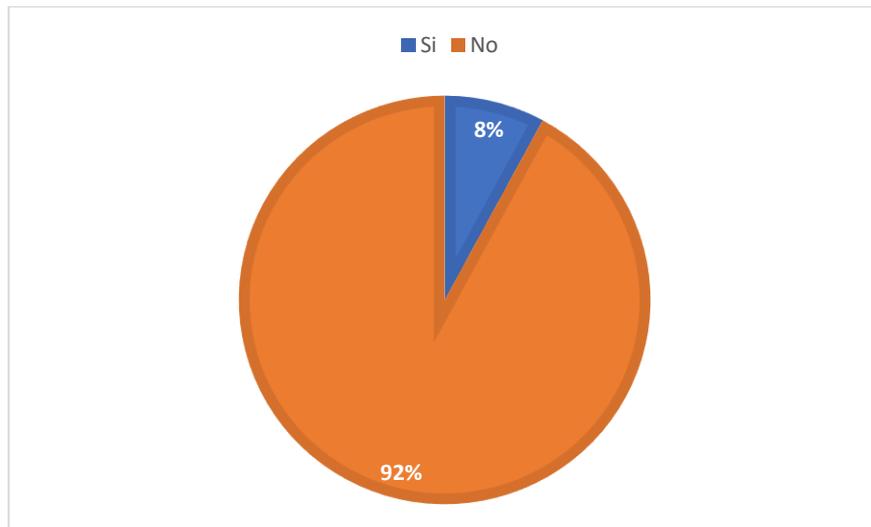
Figura 70: Bullying por parte de compañeros



Fuente: Grupo de trabajo

De acuerdo a los datos obtenidos a partir de la encuesta la mayoría de los estudiantes no han sufrido bullying por parte de sus compañeros, lo que corresponde a un 91% y una mínima parte que es un 9% de los alumnos han sido víctima de discriminación por parte de sus compañeros, lo cual puede influir en la decisión de abandonar los estudios. Por tal motivo, el bullying puede ser de tipo física o verbal que se repiten, logrando que la víctima se angustie y se encuentre en poder del acosador, con el uso de tecnologías existe más formas de bullying (Unicef, 2010).

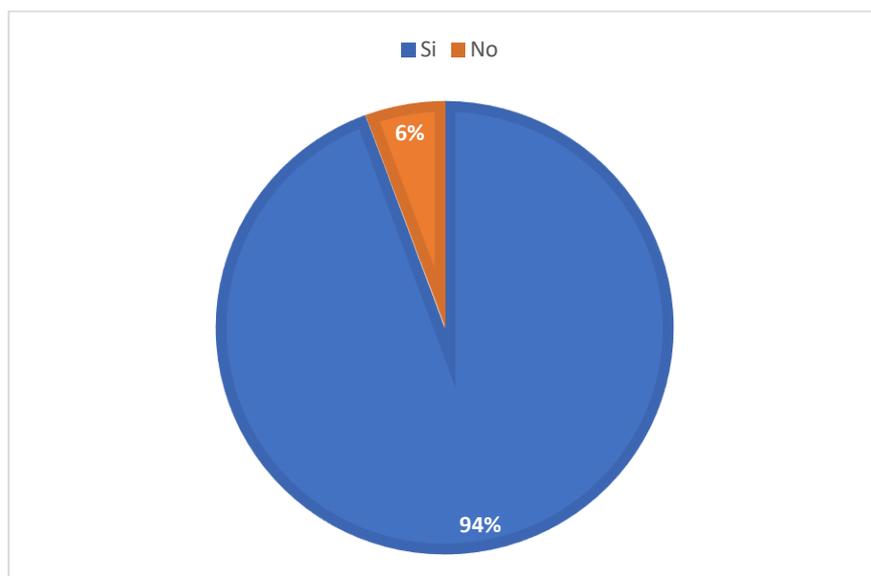
38.- ¿Ha sido víctima de discriminación o bullying por parte de sus docentes?

Figura 71: Bullying por parte de docentes

Fuente: Grupo de trabajo

La mayoría de los encuestados correspondiente a un 92% no han sido víctimas de discriminación por parte de los docentes en el entorno en el que se encuentran, mientras que un 8% de los estudiantes afirman que han sufrido de bullying de los docentes. De tal manera, el bullying es considerado como abuso de poder, violencia, agresión sobre una persona que se puede dar de manera periódica en tiempo y espacio (Salas Picón, 2015).

39. ¿Las expectativas profesionales que tiene planeadas a futuro son aceptadas por su familia?

Figura 72: Expectativas profesionales planeadas

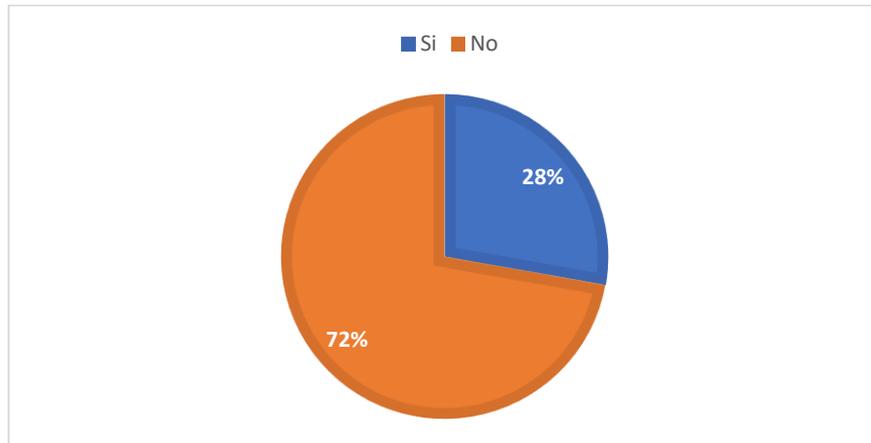
Fuente: Grupo de trabajo

La mayor parte de los estudiantes afirman que los planes profesionales o las metas trazadas que tienen planeadas a futuro son aceptados por su familia, lo cual es equivalente a un 94% de

respuestas y una pequeña parte del 6% de los planes no son aprobados por la familia del alumno por tal razón dichos estudiantes pueden estar en riesgo de desertar.

40. ¿Ha pensado usted en algún momento en abandonar la universidad?

Figura 73: Abandonar la universidad

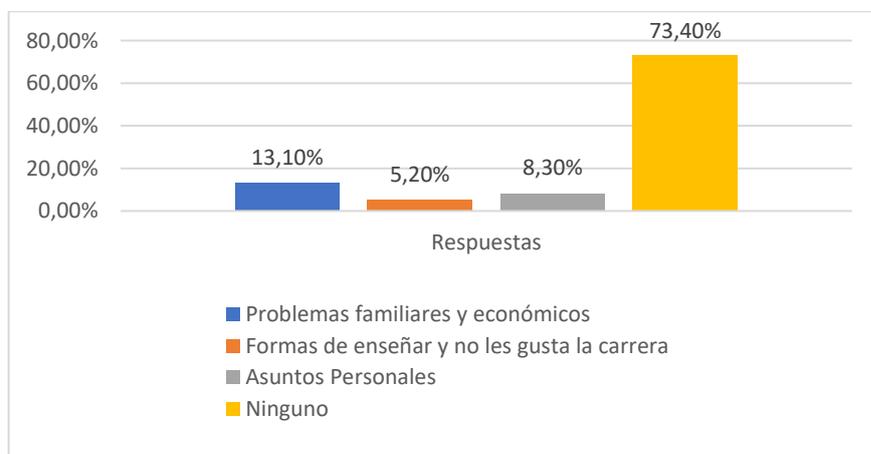


Fuente: Grupo de trabajo

La deserción se presenta en todos los niveles de universidad, pues son los estudiantes de más bajos recursos que renuncian a tener un título, la mayoría de los abandonos se da por circunstancias que obligan a dejar sus estudios (Salcedo, Ramírez, & Díaz, 2017). Según los datos de la encuesta un 72% de los estudiantes no han pensado en abandonar sus estudios universitarios pese a las dificultades que se les presente y un 28% de los alumnos han considerado en dejar la universidad.

41. Si su respuesta es sí, indique las razones

Figura 74: Razones para dejar la universidad

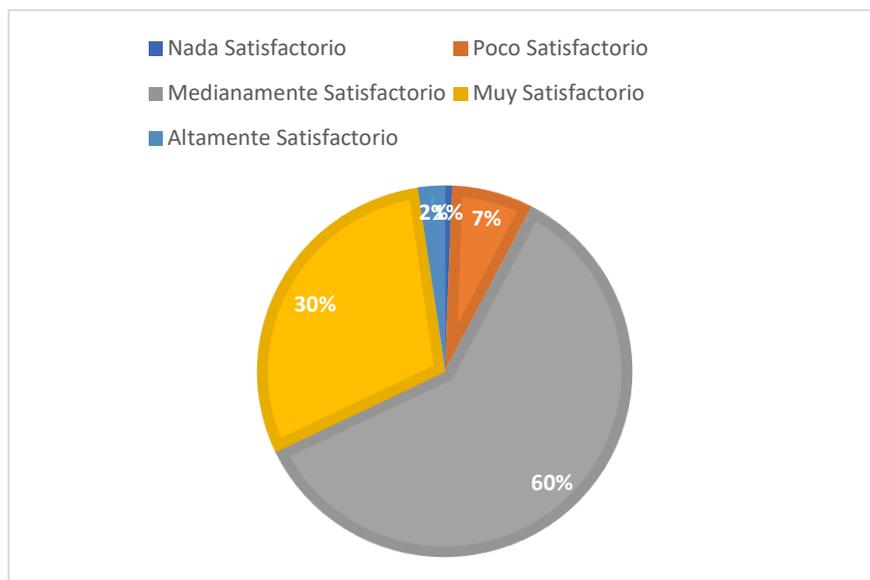


Fuente: Grupo de trabajo

En base a la pregunta 40 se puede determinar que existen diversos motivos y causas que han hecho pensar a los estudiantes en dejar sus estudios académicos, siendo la forma de enseñar y porque no les gusta la carrera el de menor porcentaje 5,2% de respuestas, un 8,3% son por razones personales que tiene cada alumno, por problemas familiares y económicos un 13,1% que representa el porcentaje mayor de las razones expuestas.

42. ¿Su rendimiento académico es satisfactorio?

Figura 75: Rendimiento académico de los estudiantes

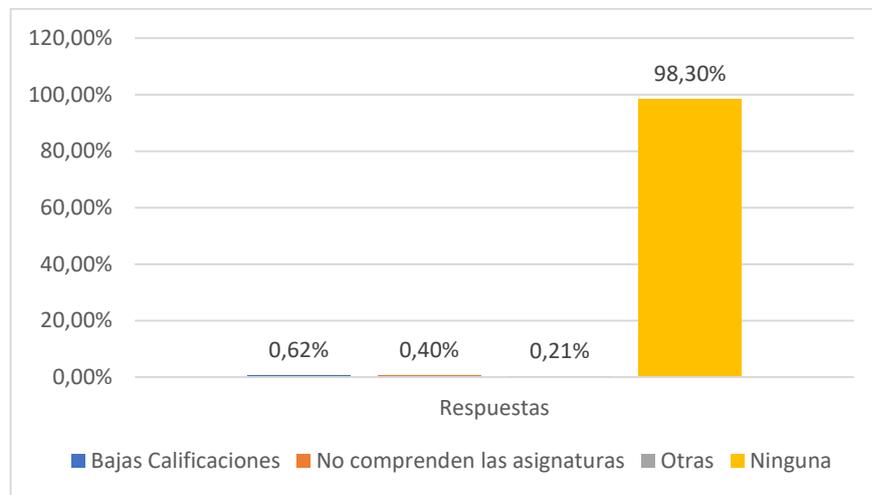


Fuente: Grupo de trabajo

El nivel académico se determina mediante los conocimientos que se demuestra en un área o materia sin importar la edad y el nivel académico, se valora de acuerdo a procesos de evaluación. Los docentes consideran variables que determinan el rendimiento académico entre ellas se encuentra las calificaciones (Navarro, 2008). Por consiguiente, los datos obtenidos muestran que un 60% de los estudiantes han manifestado que tienen un rendimiento académico medianamente satisfactorio y un 30% muy satisfactorio, estos dos porcentajes son considerados altos, mientras que los alumnos con notas altas corresponden al 2% lo que conlleva a un rendimiento académico altamente satisfactorio.

43. Si su respuesta es nada satisfactorio, indique las razones

Figura 76: Razones de bajo rendimiento académico

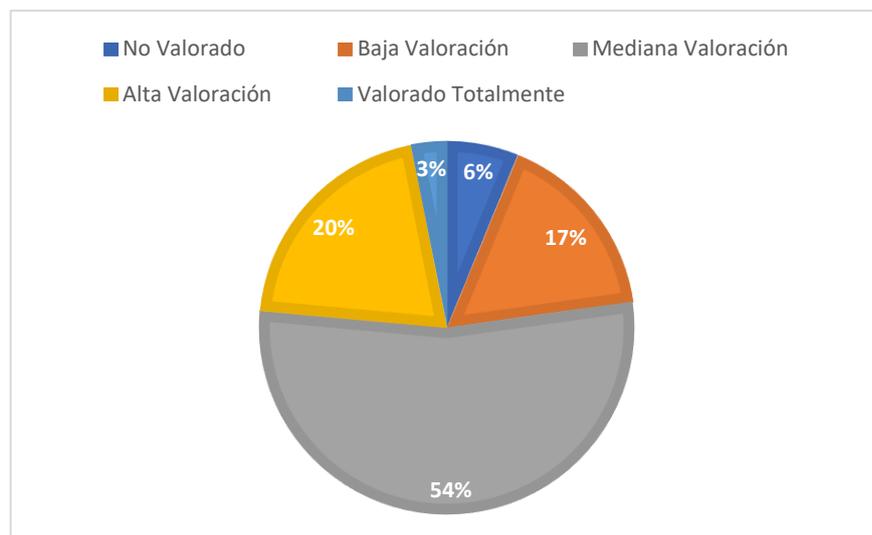


Fuente: Grupo de trabajo

De acuerdo a la pregunta 42 existen pocas respuestas con un rendimiento académico poco satisfactorio, los motivos expuestos por los estudiantes son las bajas calificaciones correspondiente a un 0,62%, el no comprender la asignatura 0,40%, además añaden otras razones que representa el 21%.

44. ¿Considera usted, que sus profesores valoran su esfuerzo y trabajo en el aula de clases?

Figura 77: Valoran el esfuerzo los docentes

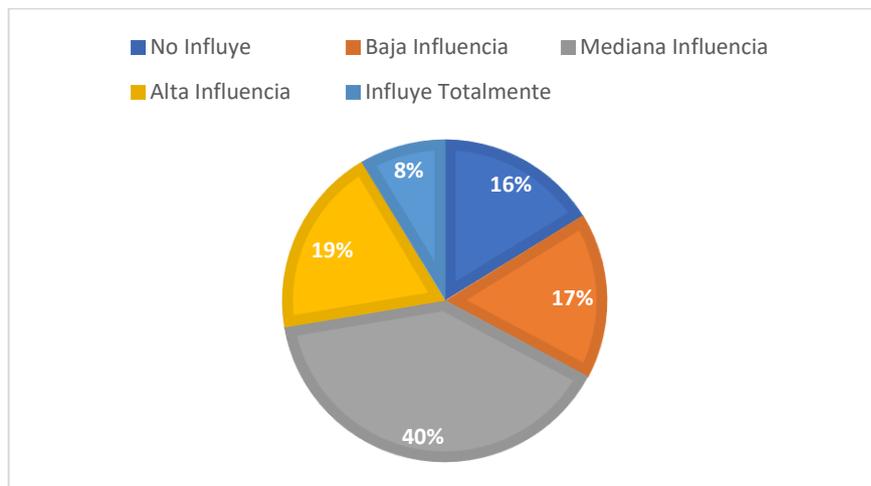


Fuente: Grupo de trabajo

Según las respuestas un alto porcentaje de estudiantes piensan que sus docentes valoran el esfuerzo que realizan en diversas actividades propuestas que se pueden dar en el aula de clase, así como fuera de la misma, esto corresponde a un 54% de una mediana valoración, el 20% considera una alta valoración, el 3% piensa que existe una baja valoración y un 6% de alumnos consideran que no valoran el esfuerzo que realizan los estudiantes.

45. ¿Considera usted, que un inadecuado sistema de tutorías influye negativamente en la decisión de los estudiantes de abandonar las aulas universitarias?

Figura 78: Inadecuado sistema de tutorías

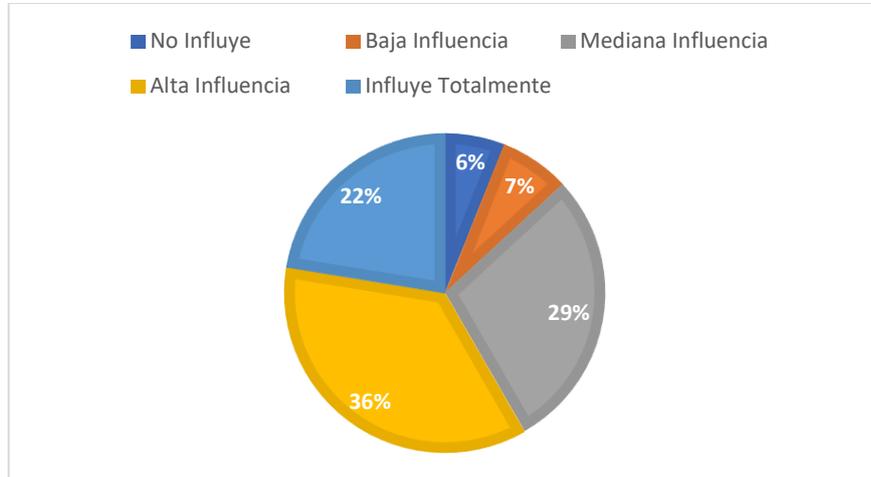


Fuente: Grupo de trabajo

De acuerdo a los datos obtenidos por los estudiantes un 40% que es equivalente a una mediana influencia manifestó que un inadecuado sistema de tutorías por parte de los docentes influye para que los alumnos decidan no continuar con sus estudios superiores y un 8% piensa que influye totalmente en la decisión de abandonar la universidad. Pues las tutorías académicas deben encontrarse en una organización que conlleva planificar, ejecutar, verificar y actuar para la toma de experiencias con el fin de obtener resultados que tengan beneficios. (Correa Peralta, Vinueza Martínez, Torres Arias, & Ponce Intriago, 2019)

46. ¿Considera usted, que los problemas familiares influyen negativamente en la decisión de abandonar la universidad?

Figura 79: Problemas familiares

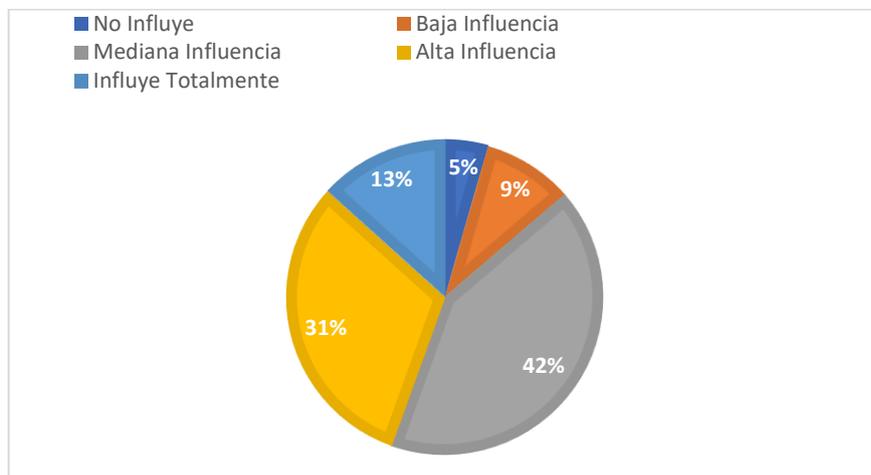


Fuente: Grupo de trabajo

Los problemas familiares afectan a la formación de los estudiantes, pues la familia es el núcleo de la sociedad en la cual se fortalece o debilitan los deseos de estudiar (Pastrana, Guitiérrez, & Mendoza, 2018) por tal motivo, de los encuestados un 36% correspondiente a una alta influencia respondieron que los problemas familiares que llega a tener un estudiante pueden afectar a continuar con sus estudios académicos, mientras que un 6% que es equivalente a la minoría de respuestas manifestó que no influye en la decisión de abandonar las aulas universitarias.

47. ¿Cuál es el grado de influencia que tiene una inadecuada formación académica del docente con la deserción estudiantil universitaria?

Figura 80: Inadecuada formación del docente

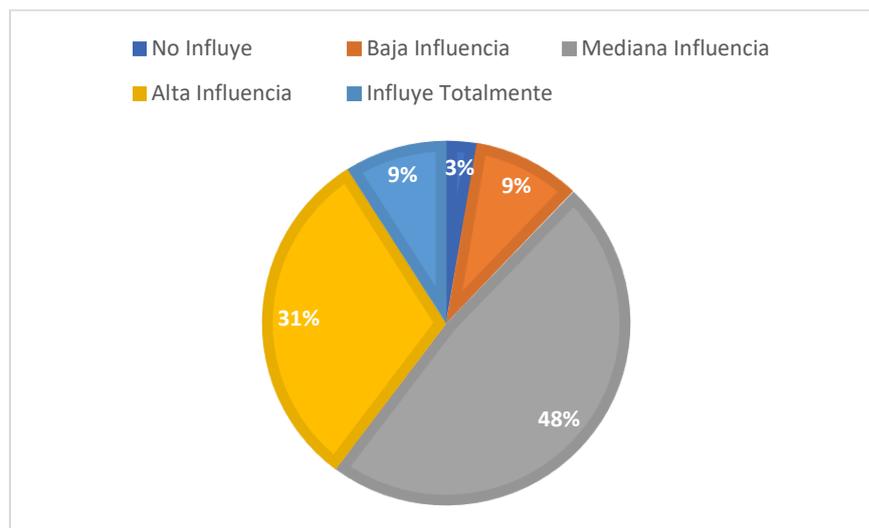


Fuente: Grupo de trabajo

La mayoría de los estudiantes correspondiente a un 42% consideran que los docentes al no contar con una adecuada formación académica influyen medianamente a que los alumnos que estén cursando los primeros ciclos académicos decidan en abandonar la universidad, pues la formación de los docentes universitarios debe alcanzar la calidad y excelencia del sistema de educación, además de que debe tener conocimientos de filosofía, psicología y pedagogía (Padilla Gómez, López Rodríguez del Rey, & Rodríguez Morales, 2015). Inclusive un 32% piensa que es de alta influencia el que un docente no se encuentre preparado para impartir su cátedra.

48. ¿Cuál es el grado de influencia que tiene una inadecuada orientación del aprendizaje con la deserción estudiantil universitaria?

Figura 81: Inadecuada orientación del aprendizaje

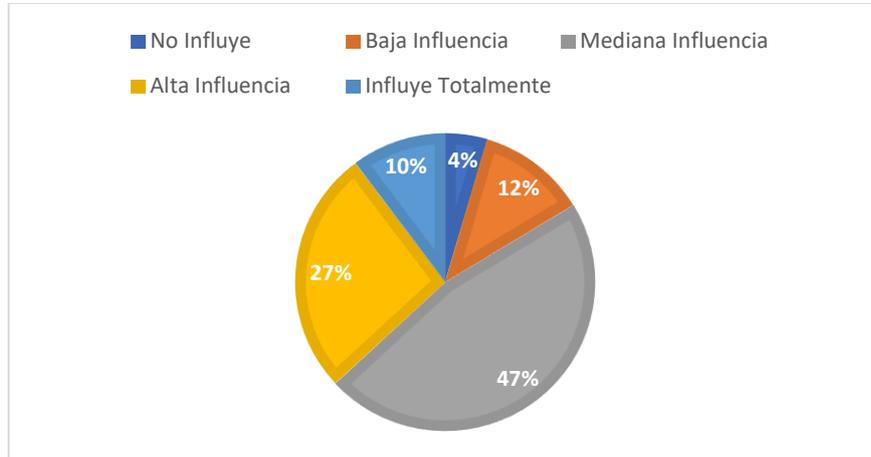


Fuente: Grupo de trabajo

La dificultad de aprendizaje están ligados a los trastornos de una lesión cerebral de tipo psicológico que impide el desempeño intelectual y la capacidad para adquirir nuevos conocimientos (Campos & Vargas, 2009). Con los datos obtenidos de los estudiantes se puede describir que un 48% de las respuestas creen que cuando no existe una orientación que este acorde al aprendizaje para los estudiantes, lo que corresponde a que sea una media influencia para los alumnos que tomen la decisión de no continuar con su carrera profesional y un 3% manifiesta que no influye ni afecta al estudiante.

49. ¿Considera usted, que un inadecuado desarrollo de las cualidades humanas del docente en el aula de clases puede influir en la deserción estudiantil?

Figura 82: Inadecuado desarrollo de cualidades humanas del docente

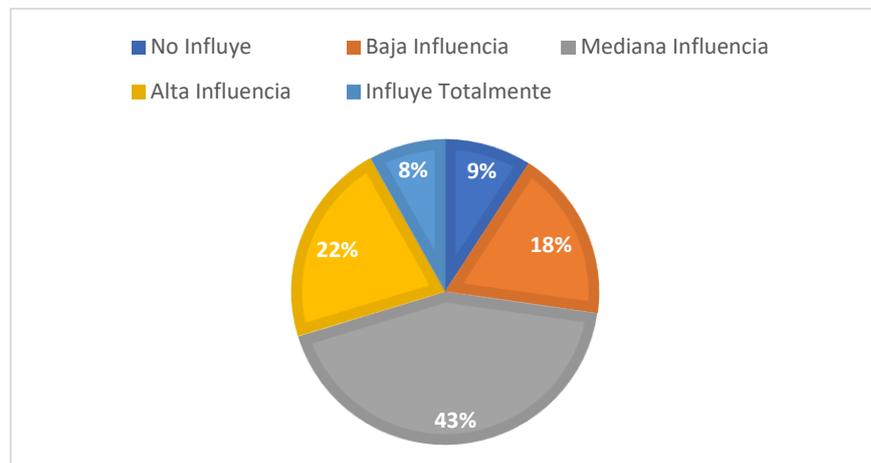


Fuente: Grupo de trabajo

La personalidad es un conjunto de elementos que establece la forma propia de una persona, en la que responde a los estímulos y circunstancias de la vida, con lo cual se determina las cualidades que se destacan (Valencia Merizalde, 2010). Por tal razón, el inadecuado desarrollo de las cualidades humanas influye en la decisión de los alumnos para abandonar los estudios universitarios, según las encuestas aplicadas la mayoría de los estudiantes manifestaron que es de mediana influencia correspondiente a un 47%, un 27% cree que es de alta influencia, un 12% que es de baja influencia, un 10% que influye totalmente y un 4% que no existe influencia.

50. ¿Considera usted, que un limitado conocimiento para la utilización de tecnologías especializadas de su carrera influye negativamente en los estudiantes de abandonar la universidad?

Figura 83: Inadecuado desarrollo de cualidades humanas del docente

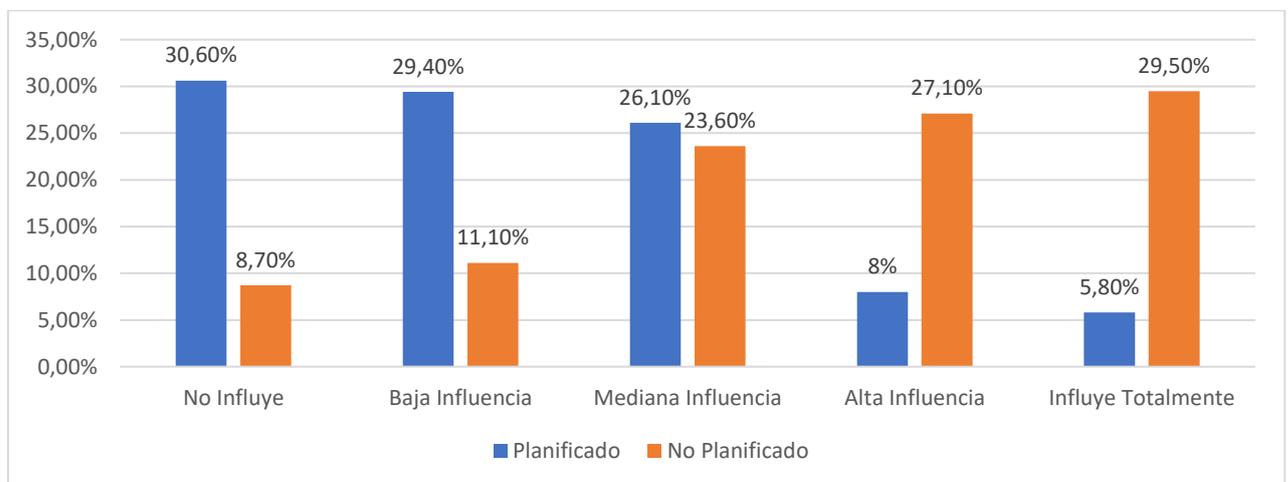


Fuente: Grupo de trabajo

La mayoría de las respuestas obtenidas, consideran que el limitado conocimiento sobre la utilización de las herramientas tecnológicas para el aprendizaje de la carrera que estudia cada uno de los estudiantes que respondió la encuesta influyen medianamente en los estudiantes para abandonar la universidad lo que corresponde a un 43%, mientras que un 9 % piensa que no influye en la decisión de desertar.

51. ¿Considera usted, que el embarazo planificado – no planificado influye en la decisión de los estudiantes de abandonar la universidad?

Figura 84: Estado de gestación del estudiante planificado - no planificado

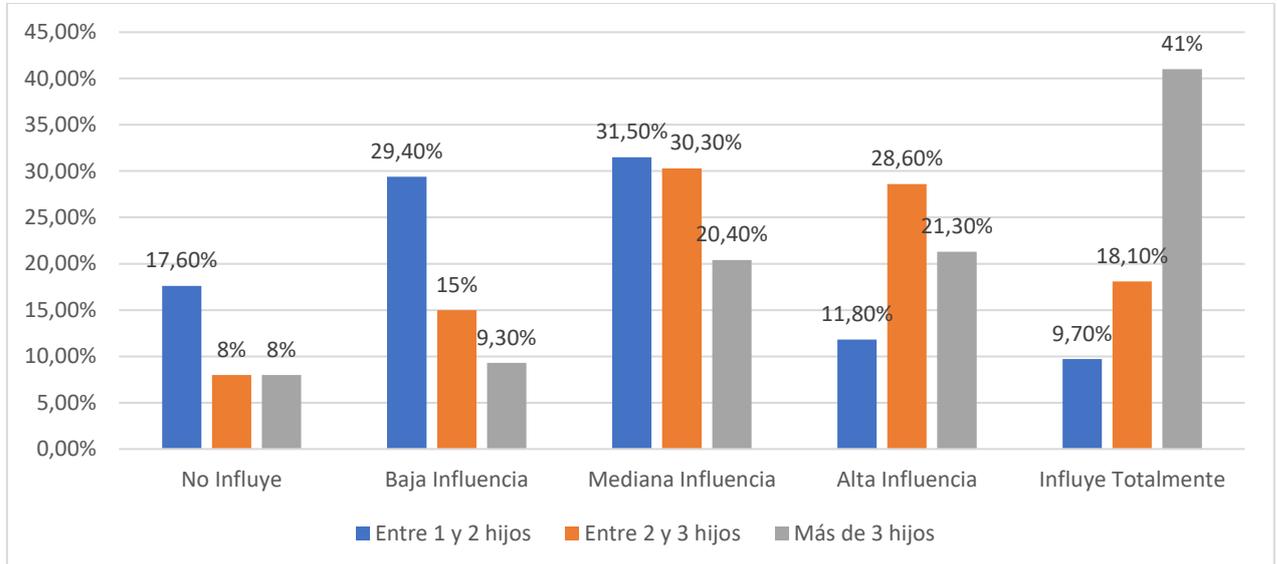


Fuente: Grupo de trabajo

Según los datos obtenidos la mayoría de los estudiantes, un 43% consideran que el embarazo planificado, no influye en la decisión de desertar mientras que un 5,8% cree que influye totalmente en dicha decisión, por el contrario, una estudiante al no planificar su embarazo influye totalmente en la decisión de abandonar sus estudios lo que corresponde a un 29,5% de las respuestas. Pues el embarazo no deseado lleva a una serie de responsabilidades, ya que a cierta edad los estudiantes no se sienten preparados para estos cambios, además de la crisis económica personal o de pareja. La consecuencia del embarazo no deseado conlleva a interrumpir sus estudios universitarios y empezar a trabajar para tener recursos económicos (K. Pérez & Orozco, 2011).

52. ¿Cuál es el nivel de influencia que tiene el número de hijos de los estudiantes con la deserción estudiantil universitaria?

Figura 85: Estado de gestación del estudiante planificado - no planificado

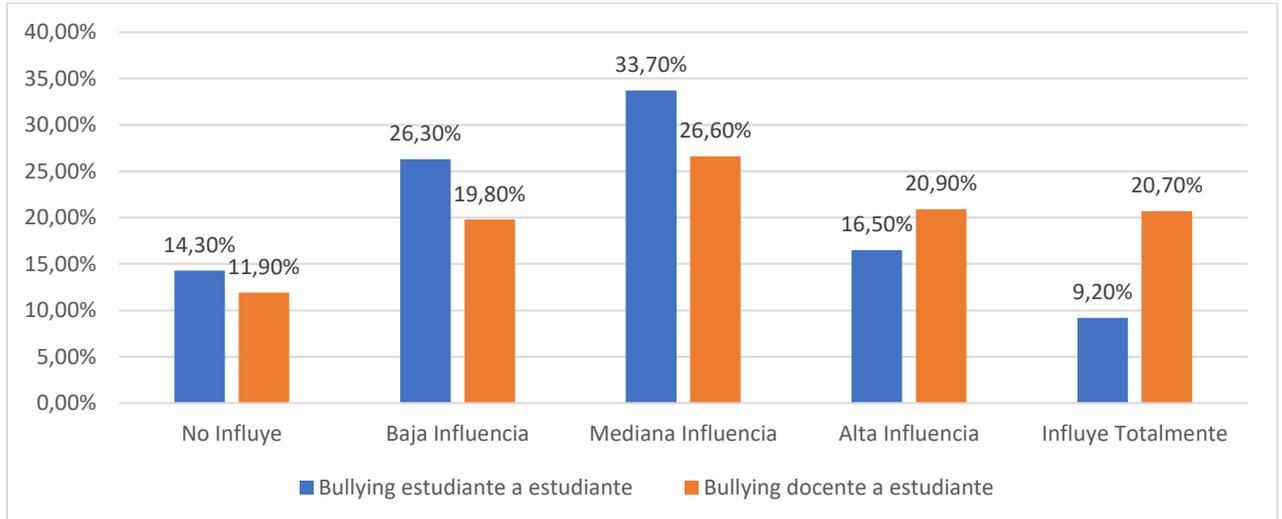


Fuente: Grupo de trabajo

Mediante la encuesta aplicada a los estudiantes se pudo constatar que un 43% de respuestas considera que influye totalmente en un alumno no continuar con su carrera profesional al tener más de 3 hijos, pues existe mayor responsabilidad, mientras que es de mediana influencia el tener entre 2 y 3 hijos lo que corresponde a un 30,3% de igual manera que entre 1 y 2 hijos que es un equivalente a 31,5% ya que en algunos casos no tienen con quien dejar a sus hijos para que los cuide mientras sus madres estudian.

53. ¿Cuál es el nivel de influencia que tienen los grupos que componen la comunidad universitaria respecto al factor BULLYING, con la deserción estudiantil universitaria?

Figura 86: Factor bullying

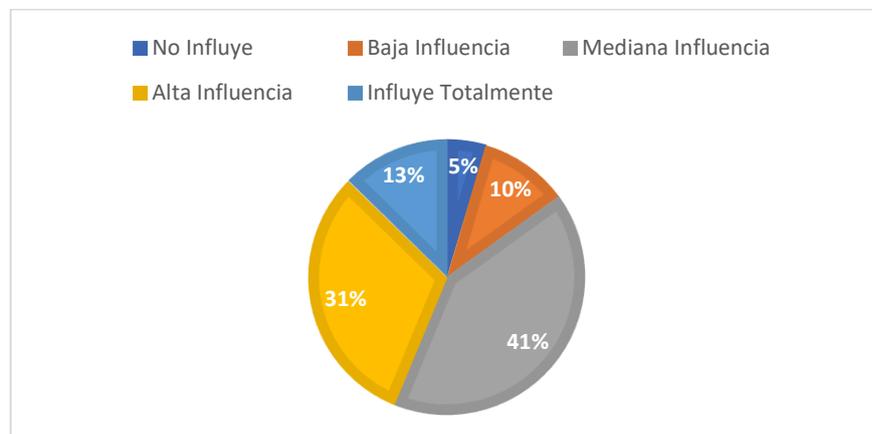


Fuente: Grupo de trabajo

De acuerdo a los datos que se obtuvieron existe un alto índice de que el bullying puede existir entre estudiante a estudiante con un porcentaje de 33,7% y bullying que puede darse entre un docente a estudiante correspondiente a un 26,6% , los mismos que son considerados como mediana influencia esto puede ser una causa para que los estudiantes abandonen la universidad, mientras que un 14,3% y 11,9% de las respuestas que se puede dar entre los grupos ya mencionados manifiestan que no tiene influencia en la decisión de desertar

54. ¿El limitado conocimiento de la asignatura por parte del estudiante influye en su necesidad de desertar?

Figura 87: Factor bullying

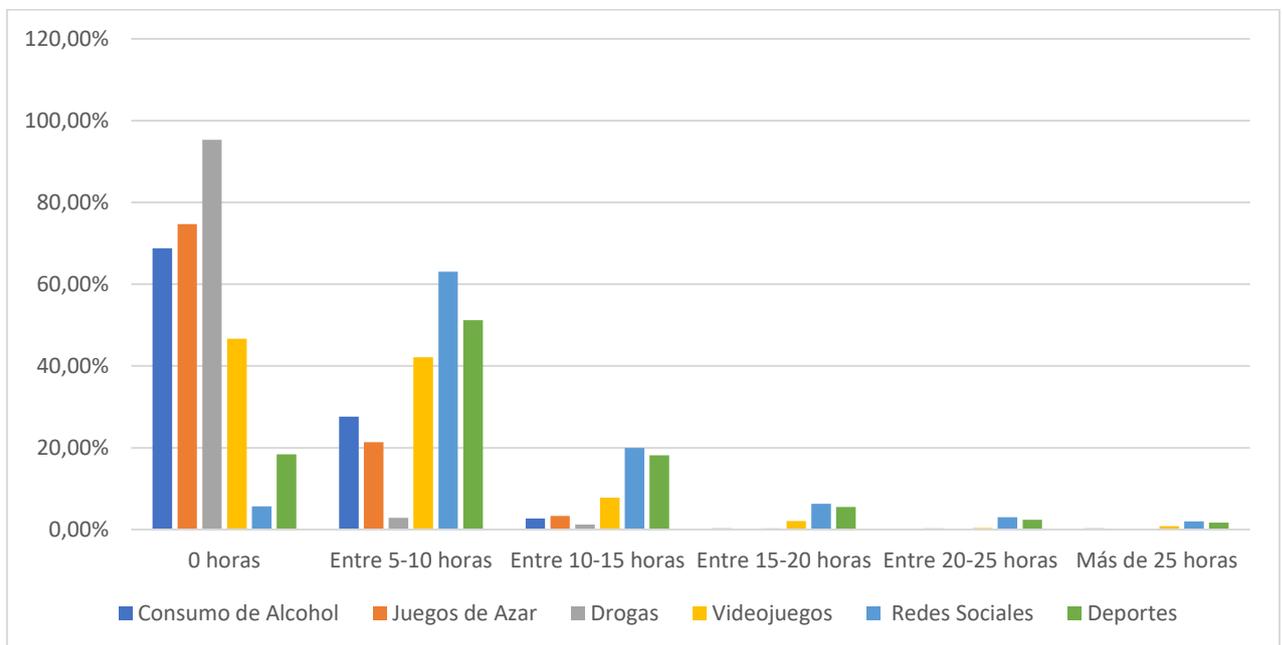


Fuente: Grupo de trabajo

El conocimiento es la forma de transmitir conocimientos propios o aquellos que se encuentren interconectados a otros conocimientos, lo que conlleva a una red que no tiene que ser la misma que la del receptor (Padilla, n.d.). Por tal motivo, un 41% de los estudiantes manifestó que es de mediana influencia el limitado conocimiento de una asignatura por parte del estudiante para que tome la decisión de no continuar con sus estudios universitarios, el 31% cree que es de alta influencia, el 13% considera que influye totalmente, el 10% manifiesta que es de baja influencia y el 5% establece que no influye en la decisión del alumno.

55. ¿Cuánto tiempo a la semana les dedica a las siguientes actividades?

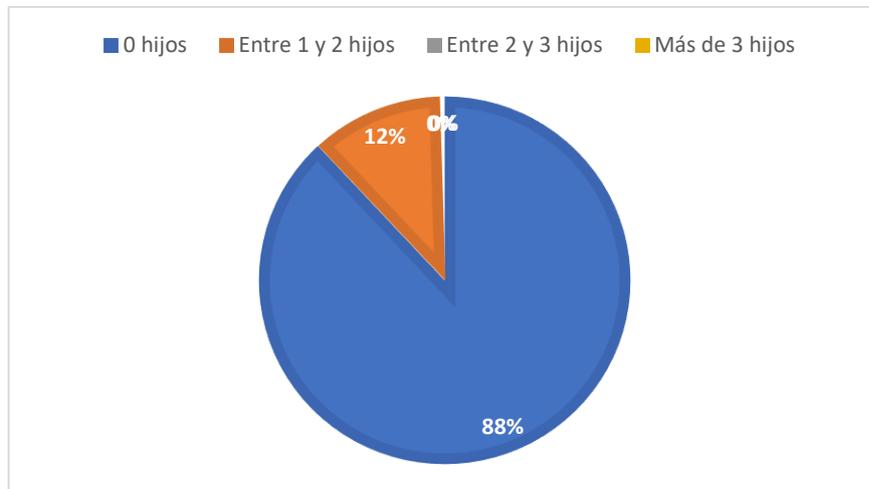
Figura 88: Actividades que se dedica el estudiante



Fuente: Grupo de trabajo

De acuerdo a los resultados obtenidos de las encuestas aplicadas, la mayoría de los alumnos no se dedican a actividades que puedan afectar a su salud así el 95% de estudiantes no consumen drogas, el 74,4% no se dedica a juegos de azar y el 68,8% no consumen bebidas alcohólicas, mientras que entre 5 a 10 horas existe un 63,1% de estudiantes que están inmersos en las redes sociales y el 42,1% se dedican a los videojuegos por otro lado existe un considerable grupo de alumnos que están dedicados al deporte con un 51,2%.

56. ¿Cuántos hijos tiene usted?

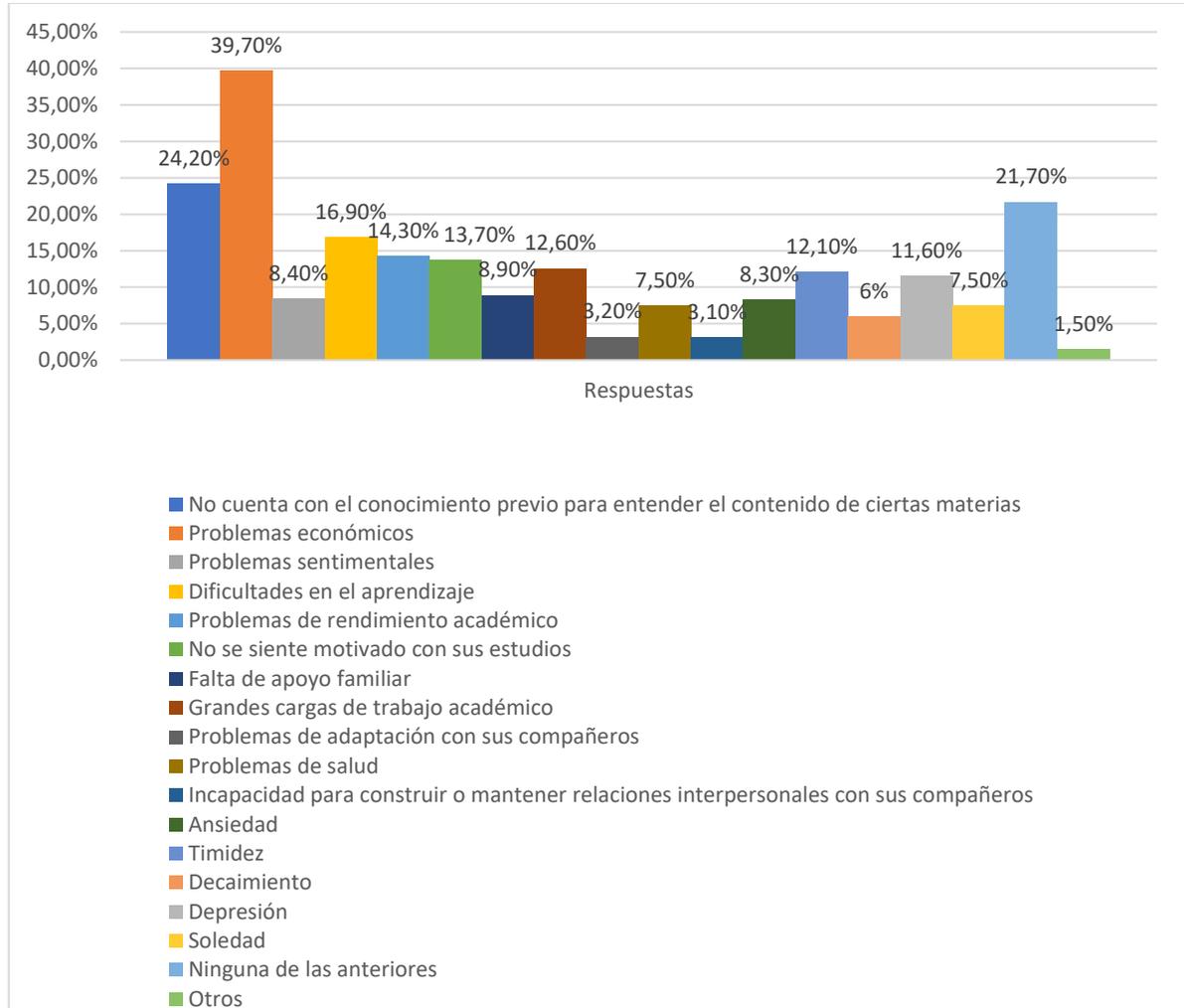
Figura 89: Actividades que se dedica el estudiante

Fuente: Grupo de trabajo

La mayoría de los estudiantes que están cursando actualmente sus estudios en el periodo académico 2019, no tienen hijos lo que indica que existe un alto porcentaje que corresponde a un 88% de estudiantes sin ningún tipo de responsabilidad de ser padres, mientras que un 12% de los alumnos manifestaron que tienen entre 1 y 2 hijos a los que deben cuidar, pero no es una limitante para continuar con sus estudios universitarios.

57. De los siguientes ítems ¿Cuál de estos se relaciona con su estado emocional actual que imposibiliten continuar con sus estudios universitarios?

Figura 90: Factores que impiden continuar sus estudios

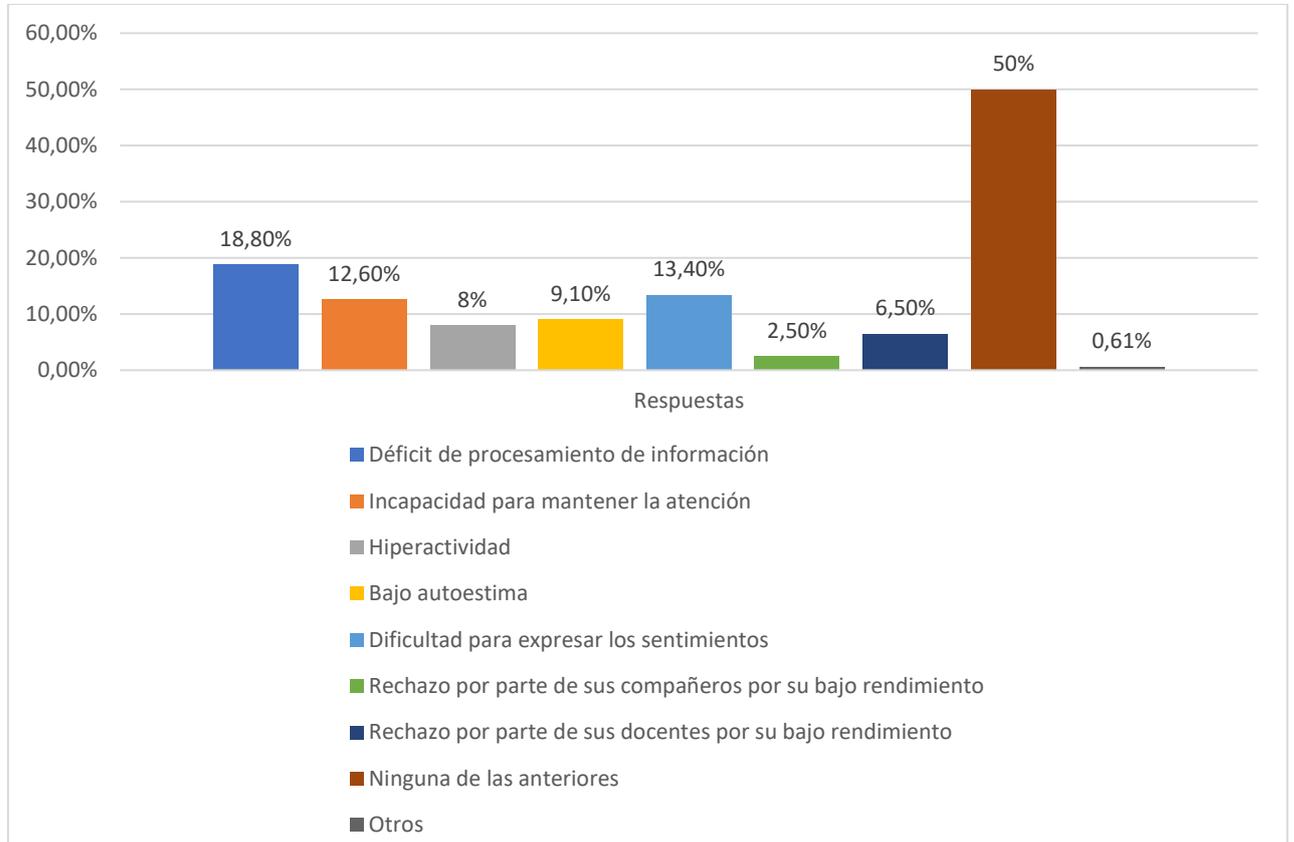


Fuente: Grupo de trabajo

Mediante las encuestas aplicadas existen diferentes factores que impiden actualmente a los estudiantes continuar con sus estudios académicos, pues se obtuvieron datos en los que los principales factores escogidos por los alumnos son el 39% correspondiente a problemas económicos, el 24,2% no cuenta con el conocimiento previo para comprender el contenido de ciertas asignaturas y un 21,7% considera que la timidez puede afectar para que abandonen la carrera que se encuentran estudiando.

58. De los siguientes factores indique ¿Cuáles se relacionan con usted en su etapa de formación académica?

Figura 91: Factores relacionados con la etapa académica

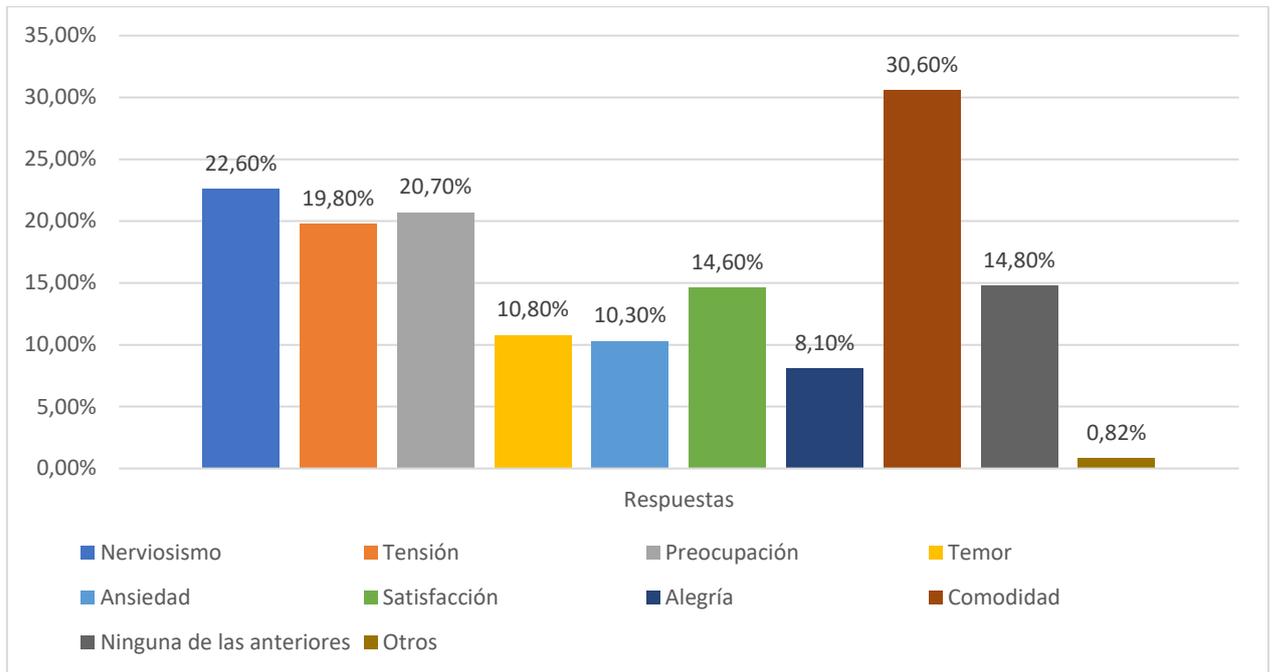


Fuente: Grupo de trabajo

De acuerdo a la encuesta aplicada, se pudo determinar que los factores que se relacionan con la etapa de formación académica, que los estudiantes consideran como factores principales son el déficit para procesar información con un 18,8% de respuestas, un 13,4% tienen problemas para expresar los sentimientos y un 12% que tiene dificultades para mantener la atención en clases, mientras que por otro lado, un alto porcentaje del 50% de alumnos consideran que de los posibles factores planteados ninguno les afecta.

59. ¿Qué estado de ánimo presenta usted en relación con el docente en el aula de clases?

Figura 92: Estado de ánimo frente al docente



Fuente: Grupo de trabajo

El estado de ánimo es la emoción que siente un individuo en un determinado momento, es la forma de estar o permanecer del mismo. Este estado de ánimo puede ser agradable o desagradable y está asociado con el humor con el que se encuentre el individuo en aquel instante (Martínez, 2018). Así que la mayoría de los estudiantes presentan un estado de ánimo de comodidad en relación con el docente en el aula de clases que corresponde a un 30,6% de respuestas, aunque existen otros estados de ánimo que tienen los alumnos al momento de encontrarse con el docente en el aula de clases como el nerviosismo con el 22,6% y el 20,7% de preocupación.