



UNIVERSIDAD TÉCNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRÍA EN SISTEMAS DE INFORMACIÓN

MODALIDAD: PROPUESTA METODOLÓGICA Y TECNOLÓGICA AVANZADA

Título:

**MODELO PARA LA DETERMINACIÓN DE FACTORES DE DESERCIÓN
ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DE COTOPAXI
UTILIZANDO MINERÍA DE DATOS.**

Trabajo de titulación previo a la obtención del título de magíster en Sistemas de
Información

Autor
Ing. Edwin Geovanny Chimba Lagla
Tutor
PhD. Mayra Susana Albán Taipe

**LATACUNGA –ECUADOR
2020**

APROBACIÓN DEL TUTOR

En mi calidad de Tutor del Trabajo de Titulación “**MODELO PARA LA DETERMINACIÓN DE FACTORES DE DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DE COTOPAXI UTILIZANDO MINERÍA DE DATOS**” presentado por el Ing. Edwin Geovanny Chimba Lagla para optar por el título Magíster en Sistemas de Información.

CERTIFICO

Que dicho trabajo de investigación ha sido revisado en todas sus partes y se considera que reúne los requisitos y méritos suficientes para ser sometido a la presentación para la valoración por parte del Tribunal de Lectores que se designe y su exposición y defensa pública.

Latacunga, 09, 24, 2020

.....

Dra. Albán Taipe Mayra Susana

0502311988

APROBACIÓN TRIBUNAL

El trabajo de Titulación: “MODELO PARA LA DETERMINACIÓN DE FACTORES DE DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DE COTOPAXI UTILIZANDO MINERÍA DE DATOS” ha sido revisado, aprobado y autorizado su impresión y empastado, previo a la obtención del título de Magíster en Sistemas de Información; el presente trabajo reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la exposición y defensa.

Latacunga, 09, 24, 2020.

.....

MSc. José Cadena
0501552798
Presidente del tribunal

.....

MSc. Karla Susana Cantuña Flores
0502305113
Lector 2

.....

PhD. Gustavo Rodríguez Bárcenas
1757001357
Lector 3

DEDICATORIA

Este trabajo de investigación está dedicado a:

A mis padres Nelson y Blanca por inculcarme el respeto, la humildad y la responsabilidad que tenemos los profesionales en nuestra preparación académica constante, a mi esposa Sandra Cevallos por su apoyo y paciencia incondicional, a mis hij@s Madeline, Amarus y Aramis, quienes se han convertido en el pilar fundamental para alcanzar este objetivo, con su ternura, amor y fuerza para no desmayar, superado así toda adversidad que se presente.

A Josué, Lolita y Mica, quienes siempre estarán en mis pensamientos y en mi corazón.

Edwin

AGRADECIMIENTO

Un agradecimos especial a la Dra. Mayra Albán por su apoyo incondicional como tutora del presente proyecto de investigación, a la Ing. Susana Benalcázar Wated por su apoyo incondicional para mi crecimiento profesional, a mi tía Digna Chimba que se ha convertido en mi segunda Madre velando siempre por mi bienestar lejos de casa, a los docentes de la Maestría por su aporte en mi formación académica, a los compañeros de la Maestría por ser unas excelentes personas y profesionales.

A la Universidad Técnica de Cotopaxi por brindarme esta oportunidad de alcanzar nuevos objetivos académicos de alto nivel.

Edwin

RESPONSABILIDAD DE AUTORÍA

Quien suscribe, declara que asume la autoría de los contenidos y los resultados obtenidos en el presente trabajo de titulación.

Latacunga, 09, 24, 2020

A handwritten signature in blue ink, appearing to read 'Edwin Geovanny Chimba Lagla', is written over a horizontal dotted line.

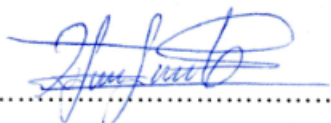
Ing. Edwin Geovanny Chimba Lagla

0502864879

RENUNCIA DE DERECHOS

Quien suscribe, cede los derechos de autoría intelectual total del presente trabajo de titulación a la Universidad Técnica de Cotopaxi.

Latacunga, 09, 24, 2020.



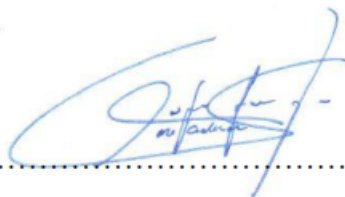
.....
Ing. Edwin Geovanny Chimba Lagla

0502864879

AVAL DEL VEEDOR

Quien suscribe, declara que el presente Trabajo de Titulación: “MODELO PARA LA DETERMINACIÓN DE FACTORES DE DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DE COTOPAXI UTILIZANDO MINERÍA DE DATOS” contiene las correcciones a las observaciones realizadas por los lectores en sesión científica del tribunal.

Latacunga, 09, 24, 2020



.....
MSc. José Augusto Cadena Moreano

CC: 0501552798

UNIVERSIDAD TÉCNICA DE COTOPAXI
DIRECCIÓN DE POSGRADO
MAESTRÍA EN SISTEMAS DE INFORMACIÓN

Título: Modelo para la determinación de factores de deserción estudiantil en la Universidad Técnica de Cotopaxi utilizando Minería de Datos.

Autor: Ing. Edwin Geovanny Chimba Lagla

Tutor: PhD. Albán Taipe Mayra Susana

RESUMEN

La educación superior en el Ecuador juega un rol muy importante en la búsqueda del desarrollo y el bienestar social, convirtiéndose así en un eje principal para el desarrollo nacional. El interés por abordar la investigación para la determinación de factores de deserción en la Universidad Técnica de Cotopaxi, se debe a que la deserción estudiantil universitaria se ha convertido en una problemática actual que afecta al estudiante, a su entorno familiar, a las universidades y a la sociedad en general. Por tanto, la determinación de factores de deserción en las universidades puede ser considerada como una estrategia clave para la toma de decisiones institucionales. Para el desarrollo de la investigación se construye un modelo teórico de deserción estudiantil universitaria a través de validación de expertos, los datos obtenidos fueron comprobados mediante Regresión Lineal con mínimos cuadrados ordinarios. Se valida el modelo teórico de deserción mediante modelos de predicción utilizando Redes Neuronales Perceptrón Multicapa y Deep Learning, mediante la metodología Knowledge Discover in Data Base (*KDD*) para proyectos de minería de datos. Los resultados indican que la técnica con mayor tasa de precisión fue el algoritmo backpropagation se obtuvo una exactitud de predicción del 98,2 %. Se concluye que los modelos propuestos son adecuados en términos de confiabilidad que se sustentan bajo procedimiento experimental.

Palabras Claves: Deserción estudiantil, factores de deserción, datos, minería de datos, modelo, redes neuronales, KDD, CRISP-CM.

**UNIVERSIDAD TÉCNICA DE COTOPAXI
DIRECCIÓN DE POSGRADO**

MAESTRÍA EN SISTEMAS DE LA INFORMACIÓN.

THEME: "Determination model of student dropout factors at Technical University of Cotopaxi using data mining."

Author: Ing. Edwin Geovanny Chimba Lagla

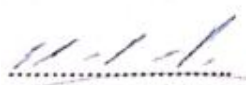
Tutor: Ph.D. Albán Taipe Mayra Susana

ABSTRACT

Superior education in Ecuador has a very important role at social welfare pursuit, thus becoming a major hub for national development. The interest in addressing research for determination of drop-out factors at Technical University of Cotopaxi is because college student desertion has become a current issue affecting their family environment, universities and society in general. Therefore, the determination of desertion factors in universities can be seen as a key strategy for institutional making-decision. For its development a theoretical model of university student desertion is built through validation by experts; the obtained data were verified by Linear Regression with ordinary least squares. The desertion theoretical model is validated using prediction models using Perceptron Multilayer and Deep Learning Neural Networks, using the Knowledge Discover in Data Base (KDD) methodology for data mining projects. The results indicate that the technique with the highest accuracy rate was backpropagation algorithm a prediction accuracy of 98.2% was obtained. It is concluded that the proposed models are suitable in terms of reliability that are supported by experimental procedure.

KEYWORDS: Student desertion, desertion factors, data, data mining, model, neural networks, KDD, CRISP-CM.

Yo, **Edison Marcelo Pacheco Pruna**, con cédula de identidad número: **0502617350** Licenciad(o) en **Ciencias de la Educación** mención **Inglés** con número de registro de la SENESCYT: 1020-12-1169234 ; **CERTIFICO** haber revisado y aprobado la traducción al idioma inglés del resumen del trabajo de investigación con el título: "Determination model of student dropout factors at Technical University of Cotopaxi using data mining." de: Edwin Geovanny Chimba Lagla, aspirante a Magister en Sistemas de la Información.


.....
Lic. Edison Marecelo Pacheco Pruna Mg.

Cc: 0502617350



**CENTRO
DE IDIOMAS**

ÍNDICE DE CONTENIDO

INTRODUCCIÓN	1
Antecedentes	3
Planteamiento del problema	3
Importancia del problema.....	6
Formulación del problema.....	8
Objetivos	8
Objetivo General	8
Objetivos Específicos	8
Justificación.....	9
Hipótesis.....	12
Metodología.....	12
CAPÍTULO I.....	15
FUNDAMENTACIÓN TEÓRICA.....	15
1.1 Deserción.....	15
1.1.1 Deserción estudiantil universitaria.....	16
1.1.2 Tipos de deserción.....	17
1.1.3 Factores de deserción	18
1.1.4 Modelos de deserción Universitaria.....	24
1.2 Proceso de extracción de conocimientos.....	27
1.2.1 Minería de datos.....	28
1.2.2 Metodología (CRISP-DM) para proyectos de minería de datos	29
1.2.3 Proceso de minería de datos (KDD).....	32
1.2.4 Fases de la minería de datos (KDD)	33
1.2.5 Modelos para la minería de datos.....	34
1.2.6 Técnicas para la minería de datos	35
1.2.7 Redes neuronales.....	36
1.2.8 Tipos de redes neuronales	37
1.2.9 Deep Learning.....	38
1.3 Conclusiones del Capítulo I	39

CAPÍTULO II	40
DETERMINACIÓN DE FACTORES DE DESERCIÓN DESDE LA PERSPECTIVA DE LOS EXPERTOS.	40
2.1 Recopilación de información.....	40
2.1.1 La Encuesta	41
2.1.2 Contacto con los encuestados.....	43
2.1.3 Ingreso de resultados de la encuesta a la plataforma Web.....	44
2.1.4 Tabulación de la información almacenada.....	45
2.1.5 Confiabilidad de los datos	46
2.1.6 Evaluación de los datos en la Escala Likert	48
2.1.7 Análística descriptiva de los datos (variables independientes).....	50
2.2 Análisis de correlación	54
2.2.1 Determinación de las variables independientes más influyentes	54
2.3 Histograma de normalidad de datos	59
2.4 Modelo de deserción estudiantil universitaria.....	62
2.4.1 Modelo teórico	62
2.4.2 Regresión lineal múltiple	65
2.4.3 Modelo conceptual	70
2.4.4 Modelo lógico	70
2.5 Conclusiones del Capítulo II	80
CAPÍTULO III.....	81
MODELOS DE PREDICCIÓN DE FACTORES DE DESERCIÓN MEDIANTE MINERÍA DE DATOS	81
3.1 Proceso experimental para determinación de factores de deserción.	81
3.1.1 Metodología	83
3.2 Fases para la implementación del proyecto.....	84
3.2.1 Definición de necesidades del cliente (comprensión del negocio).....	84
3.2.2 Estudio y comprensión de los datos.....	85
3.2.2.1 Identificación de los datos.....	85
3.2.2.2 Recolección de datos.....	85

3.2.2.3	Limpieza y transformación de los datos.....	87
3.2.3	Análisis de los datos y selección de características	89
3.2.4	Modelado	93
3.2.4.1	Selección del Algoritmo para los modelos de redes neuronales	94
3.2.4.2	Generación del Modelo	96
3.2.5	Evaluación	103
3.2.5.1	Cross-Validation del modelo de deserción estudiantil.....	106
3.2.5.2	Especificación del modelo	107
3.3	Conclusiones del Capítulo III.....	111
CONCLUSIONES GENERALES:.....		112
RECOMENDACIONES:.....		113
BIBLIOGRAFÍA		114
ANEXOS		123
Anexo 1		123
Anexo 2		129
Anexo 3		130

ÍNDICE DE TABLAS

Tabla 1: Variables personales, institucionales y circunstanciales para la retención universitaria.....	23
Tabla 2: Secciones y preguntas del cuestionario.....	42
Tabla 3: Población objetiva, para el diseño muestral.....	42
Tabla 4: Prueba de confiabilidad del instrumento de investigación	46
Tabla 5: Valores de la escala de likert.	48
Tabla 6: Valores del análisis descriptivo de las variables padres.	51
Tabla 7: Valores del análisis descriptivo de las variables hijos.	52
Tabla 8: Valores de la media y desviación estándar de las variables padres.	53
Tabla 9: Valores de la media y desviación estándar de las variables hijos.	53
Tabla 10: Variables Independientes y dependiente para la deserción estudiantil en la U.T.C.....	63
Tabla 11: Regresión lineal.	65
Tabla 12: Modelo de regresión lineal ajustado.	67
Tabla 13: Factores más influyentes del modelo ajustado de deserción estudiantil.....	71
Tabla 14: Ranking variables padres de deserción estudiantil.	91
Tabla 15: Ranking variables hijos de deserción estudiantil.	92
Tabla 16: Red neuronal – clasificación de la muestra - factores padres de deserción.....	97
Tabla 17: Red neuronal – clasificación de la muestra - factores hijos de deserción.....	97
Tabla 18: Red neuronal – resumen del modelo - factores padres de deserción	99
Tabla 19: Red neuronal – resumen del modelo - factores hijos de deserción.....	99
Tabla 20: Deep Learning – clasificación de la muestra - factores padres.....	100
Tabla 21: Deep Learning – clasificación de la muestra - factores hijos.	101
Tabla 22: Cross-validation – factores padres e hijos.	106
Tabla 23: Precisión del modelo para los factores padre e hijos.	106

ÍNDICE DE GRÁFICOS

Gráfico 1: Modelo deserción académica de Tinto. [50].	25
Gráfico 2: Modelo de Fishbein y Ajzen. [50].	26
Gráfico 3: Modelo de Ethington. [50].	27
Gráfico 4: Encuesta de uso de metodologías en proyectos de minería de datos. [60].	29
Gráfico 5: Modelo CRISP-DM. [62].	30
Gráfico 6: El proceso de descubrimiento de conocimientos en bases de datos (KDD). Tomado de Maimon y Rokac [66].	33
Gráfico 7: Proceso de contacto con los encuestados.	43
Gráfico 8: Diagrama de la aplicación web desarrollada para ingresar los resultados.	44
Gráfico 9: Encuesta publicada en la web.	44
Gráfico 10: Alfa de Cronbach variables padres.	47
Gráfico 11: Alfa de Cronbach variables hijos.	47
Gráfico 12: Distribución de frecuencia para los totales de datos recopilados.	49
Gráfico 13: Distribución de frecuencia para las variables padres de deserción estudiantil.	49
Gráfico 14: Distribución de frecuencia para las variables hijos, deserción estudiantil universitaria.	49
Gráfico 15: Código del proceso descriptivo de las variables padres.	50
Gráfico 16: Código de análisis descriptivo de las variables hijos.	51
Gráfico 17: Análisis de correlación de las variables padres de deserción.	54
Gráfico 18: Factores de deserción padres más influyentes.	55
Gráfico 19: Análisis de correlación de las variables Hijos.	56
Gráfico 20: Factores de deserción hijos más influyentes.	57
Gráfico 21: Correlación de variables padres.	57
Gráfico 22: Correlación de variables Hijos.	58
Gráfico 23: Histogramas de las variables padres de deserción estudiantil.	60
Gráfico 24: Histogramas de las variables hijos de deserción estudiantil.	61
Gráfico 25: Resultados del modelo de regresión lineal.	66
Gráfico 26: Representación gráfica del modelo original de regresión lineal.	66

Gráfico 27: Visualización de los resultados del modelo ajustado de regresión lineal.	67
Gráfico 28: Resultados del modelo ajustado de regresión lineal.	68
Gráfico 29: Ajuste del modelo.	68
Gráfico 30: Factores más significativos del modelo.	69
Gráfico 31: Comprobación de supuestos del modelo de regresión lineal.	69
Gráfico 32: Modelo conceptual de deserción estudiantil para la U.T.C.	70
Gráfico 33: Proceso para la determinación de los modelos de predicción.	82
Gráfico 34: Proceso para determinar factores de deserción.	83
Gráfico 35: Encuesta para expertos.	86
Gráfico 36: Dataset inicial - opiniones expertos.	86
Gráfico 37: Limpieza y transformación de los datos.	87
Gráfico 38: Datos recolectados.	88
Gráfico 39: Total de factores de deserción.	88
Gráfico 40: Matriz de variables padres procesados.	89
Gráfico 41: Matriz de variables hijos procesados.	89
Gráfico 42: Visualización general de variables padres de deserción.	90
Gráfico 43: Visualización general de variables hijos de deserción.	90
Gráfico 44: Esquema general de una neurona “backpropagation”.	94
Gráfico 45: Algoritmo Backpropagation.	94
Gráfico 46: Función Sigmoide.	95
Gráfico 47: Estructura del perceptrón multicapa, utilizando como función de salida Sigmoide.	96
Gráfico 48: Deep learning, entrenamiento de factores padres de deserción.	102
Gráfico 49: Deep learning, entrenamiento de factores hijos de deserción.	102
Gráfico 50: Error cuadrático – factores padres de deserción.	103
Gráfico 51: Error cuadrático – factores hijos de deserción.	104
Gráfico 52: Nivel de exactitud, factores padres de deserción.	104
Gráfico 53: Nivel de exactitud, factores hijos de deserción.	105
Gráfico 54: Evaluación del modelo deep learning, factores padres de deserción.	105
Gráfico 55: Evaluación del modelo deep learning, factores hijos de deserción.	105

Gráfico 56: Modelo de deserción – factores padres de deserción.....	107
Gráfico 57: Modelo de deserción – factores hijos de deserción.	107
Gráfico 58: Red neuronal - factores padres de deserción.	108
Gráfico 59: Red neuronal - factores hijos de deserción.	109
Gráfico 60: Pseudoprobabilidad predicha de los factores padres e hijos.....	109
Gráfico 61: Precisión de la predicción de los modelos de predicción	110
Gráfico 62: Modelo deep learning, factores padres de deserción.	110
Gráfico 63: Modelo deep learning, factores hijos de deserción.....	110
Gráfico 64: Si trabaja en una institución Pública.....	123
Gráfico 65: Experiencia con deserción estudiantil universitaria.....	123
Gráfico 66: Años de experiencia con deserción estudiantil universitaria	124
Gráfico 67: En que campus desarrolla sus actividades.	124
Gráfico 68: Función que ocupa en la Institución.	125
Gráfico 69: Institución Pública o Privada.	125
Gráfico 70: Profesión de la muestra encuestada.	126
Gráfico 71: Género de los encuestados.	126
Gráfico 72: Rango de edad de los encuestados.	127
Gráfico 73: Distribución normal de la edad y años de experiencia.	127
Gráfico 74: Distribución normal de las edades de expertos en los campus de la Universidad Técnica de Cotopaxi.	128
Gráfico 75: Modelo de deserción UTC - factores padres de deserción estudiantil... ..	129
Gráfico 76: Modelo de deserción UTC - factores hijos de deserción estudiantil.	129
Gráfico 77: Código de cross_validation - factores padres de deserción estudiantil.	130
Gráfico 78: Código de cross_validation - factores hijos de deserción estudiantil.	130

INTRODUCCIÓN

La deserción estudiantil es conocida también como el abandono de un centro educativo por parte de los estudiantes a consecuencia de diversos factores tales como: económicos, sociales, emocionales, académicos entre otros. Castaño, Gallón, Gómez y Vásquez [1] sostiene que la deserción estudiantil es un fenómeno de la educación universitaria a nivel mundial, relacionada con el hecho de retiro de los estudiantes de las aulas o del sistema educativo de una institución de nivel superior.

Según, Patiño y Cardona [2] determina que la deserción estudiantil universitaria se ha convertido en un fenómeno complejo y de múltiples impactos negativos, tanto para el estudiante como para la propia universidad, así como también, para la sociedad. Los estudios sugieren niveles diferenciados de la deserción que son afectados por los modelos formativos como: el costo financiero y el tipo de programa, las condiciones biográficas y del entorno social del estudiante, así como el valor de la educación. Las universidades responden al fenómeno como institución; sin embargo, el fenómeno va más allá de las aulas y se ha constituido en un problema del sistema educativo que requiere políticas articuladas para mantener estrategias sistemáticas que enfrenten la deserción universitaria.

Para Torres [3] la deserción se ha convertido en un problema de tipo social, donde se debe identificar necesariamente al menos una causa, la revisión de la literatura ha permitido identificar que la deserción universitaria tiene varios orígenes, enfocados

en algunas ocasiones con la relación con el docente, tales como su formación, la manera en que imparte las clases a diario, la motivación, etc. En el sistema educativo la carencia de infraestructura tecnológica, limitaciones en el empoderamiento de las necesidades de estudiante; desconociendo de las problemáticas sociales y culturales, las vivencias personales, los problemas académicos que tienen los estudiantes. Respecto al ámbito social y culturales se relaciona factores como ambientes familiares patológicos y atrasados, la pobreza, analfabetismo de los padres, la marginalidad, la necesidad de laborar en edades tempranas, entre otros factores; y por último factores relacionados con el estudiante como limitaciones en el aprovechamiento de las oportunidades, su inteligencia, su autoconcepto y su forma de aprender, el no identificarse con la carrera que está siguiendo. Todos estos factores mencionados podrían considerarse como origen del fracaso o deserción estudiantil.

La mayoría de las investigaciones analizadas atribuyen que la deserción estudiantil es un problema social, este se agrava más al hablar en términos de oportunidades educativas, ya que el estudiante que deserta en sus estudios deja un puesto que pudo haber sido ocupado por otro estudiante, que a lo mejor hubiese concluido la carrera universitaria. También, tiene efectos negativos en el ámbito económico para estado en el caso de las Universidades públicas, que demanda de un presupuesto para la ejecución de las actividades académicas, administrativas y de gestión.

De acuerdo con, Páramo y Correa [4] es posible mitigar la deserción estudiantil universitaria con la implementación de estrategias y políticas públicas a favor de la educación, estas deben ser planteadas desde la comunidad académica, de tal forma que conduzcan a determinar posibles desertores y promuevan la disminución de las tasas de deserción.

Antecedentes

Para el presente trabajo de titulación es necesario mencionar, que la línea de investigación es Tecnologías de Información y Comunicación, la sublínea de investigación es Inteligencia Artificial e Inteligencia de Negocios para toma de decisiones. La investigación forma parte del proyecto de investigación Modelamiento de Algoritmos para Sistemas de Información de la Carrera Sistemas de Información de la Universidad Técnica de Cotopaxi.

Planteamiento del problema

Según Torres [3] la deserción estudiantil es un tópico que ha venido ganado terreno a partir de la década del sesenta del siglo XX, se ha visto relacionado en diferentes campos como la investigación, ámbitos socioeconómico, sociológico y en el sistema educativo. Así es como, sobre la base de las diferentes interpretaciones sobre el tema se ha generado diversas investigaciones que ha arrojado resultados y varias formas de comprender la deserción estudiantil.

Como sostiene Seidman [5] autores de talla de Vicent Tinto, Wideman, Terenzini y Astin han venido estudiado la problemática de la deserción estudiantil universitaria desde hace ochenta años aproximadamente, los mismos que han investigado y construido teorías en las que se sustentan varias investigaciones en la actualidad; lo que hace un llamado a las instituciones educativas a revisar su filosofía en la enseñanza, el ambiente y el compromiso para con los estudiantes.

Por otro lado, Cevallos [6] manifiesta que con el objetivo de incrementar el porcentaje de graduados en los centros de educación superior, los investigadores, académicos y universidades han realizado un sin número de investigaciones, trabajos científicos y experimentos, a pesar de estas iniciativas la meta no ha sido fácil de cumplir. De hecho, en Estados Unidos un país que ha tenido un gran desarrollo en el

sistema educativo superior aún continúa con limitaciones en la disminución de las tasas de deserción en las universidades.

Para Vivas [7] la combinación del estudio y trabajo o simplemente la necesidad de trabajar se considera un factor social determinante para la deserción, también considera que existe otras causas que influye en el abandono universitario como es la transición de un nivel secundario al nivel superior, hace referencia a que el estudiante tiene diferentes perspectivas sobre la universidad, en muchos de los casos estas no se cumplen a cabalidad, ya sea por el desconocimiento del medio universitario o por experiencias propias haciendo que esta problemática aumente.

Giraldo, Abad y Díaz [8] manifiesta que las Universidades tanto públicas y privadas que como actores principales el sistema educativo son las responsables de velar por la permanencia de los estudiantes en sus instituciones, al no cumplir esta exigencia solo queda en evidencia su capacidad de gestión universitaria, convirtiéndose así, en una decadente estructura organizacional de baja gestión, agilidad, eficiencia y eficacia que no puede cumplir con la sociedad. Para Clark [9] las universidades se encuentran en una encrucijada y se ven superadas por la alta tasa de deserción sin tener capacidad de respuesta institucional. La deserción estudiantil en el nivel de educación superior, más se atribuye a una decisión individual que a un problema social y educativo, denominado como la interrupción de los estudios técnicos o profesionales antes de obtener el título universitario para el cual el estudiante se matriculó [10].

Revisando el análisis realizado por Torres [3] donde la deserción estudiantil es considerando como “un fenómeno de múltiples comprensiones”, en la actualidad se ha visto relacionada por medio de diferentes discursos enfocados en el campo de la educación. Es, necesario indagar y recopilar información cuya finalidad sea acentuar cada vez más el conocer las causas de deserción estudiantil

La necesidad de elevar el número de graduados en las universidades ha sido reconocida incluso por gobiernos de países desarrollados como el de Estados Unidos; prueba de ello, en 2009, el presidente norteamericano Barack Obama determinó como objetivo fundamental de su plan de gobierno elevar el porcentaje de graduados en las universidades [11].

Tinto [12] explica el fenómeno de la deserción desde el ámbito teórico sustentado en modelos estadísticos basados en los enfoques que trabajan los factores asociados a los estudiantes en el campo individual y social, factores socioeconómicos e institucionales y la interacción de los mismos. Permitiendo así afirmar que la deserción estudiantil ha sido tratada desde perspectivas cuantitativas. La literatura ha permitido identificar artículos que conforman la fuente documental de esta investigación, se han podido reconocer diferentes formas de identificar los aspectos de deserción estudiantil, obteniendo una amplia lista relacionada con diversas realidades, por ejemplo, la deserción, la no titulación, la repitencia, el bajo rendimiento académico, la no comprensión de los temas, la exclusión, entre otros. Dando origen a una producción social compuesta por dimensiones políticas, sociales, culturales, pedagógicas, personales, cognitivas, psicológicas e Institucionales; enmarcadas en un contexto de índole familiar e individual [3].

Es importante también señalar que el estudio del abandono o deserción prematuro ha sido ampliamente reconocido como un problema que cada año se agrava más, especialmente en el nivel universitario. La educación superior en la actualidad está enfrentando dificultades comunes entre ellas, como es el incremento de las tasas de deserción y la disminución de las tasas de graduaciones. En virtud del conocimiento de las causas negativas que genera la deserción en las universidades se considera importante analizar el contexto de las causales de abandono estudiantil partiendo desde la expectativa de la percepción de expertos en la Universidad Técnica de Cotopaxi.

Importancia del problema

De acuerdo con lo expuesto por, Tinto Vincent este indicador será evaluado en los procesos institucionales, la deserción para el individuo es el fracaso para completar un curso o alcanzar una meta deseada; mientras, para la institución también es un fracaso por no ayudar al estudiante a conseguir su meta. La deserción es considerada como una problemática que aborda la mayoría de las Instituciones de Educación Superior de Latinoamérica; diversas publicaciones revelan altas cifras de deserción como consecuencia de factores socioeconómicos y educativos que se asemejan por la falta de oportunidades [13]. Además, el Centro de Investigaciones Económicas CIE (2015) indica que “las principales razones para la deserción estudiantil se encuentran en las causas económicas, familiares, personales, pedagógicas, salud, geográficas” [14]. Las universidades pierden entre 25% y 35% de su matrícula, sumando aquellos porcentajes de estudiantes que dejan estas instituciones antes de titularse [15].

La OCDE “Organización para la Cooperación y el Desarrollo Económicos” formada por 34 países miembros sostiene que un 30% de los universitarios abandonan sus estudios. Teniendo en cuenta que en países como Hungría, Suecia o Estados Unidos las tasas de deserción son de aproximadamente el 40%, mientras que en países como Australia, Dinamarca o Japón las tasas de deserción son de menos del 25% [16]. También menciona que la deserción total universitaria de Latinoamérica con un porcentaje del 57% de los estudiantes de los países como Chile y México que forman parte de la OCDE.

El estudio realizado por Viteri y Uquillas [17], tuvo como objetivos determinar las causas que han generado el abandono de los estudios universitarios; se concluye que, los factores académicos y personales son determinantes para que un estudiante decida abandonar los estudios en la universidad. Vincent Tinto [18] presenta la teoría de retención basado en los procesos institucionales, los mismos que tiene que ser evaluado, puesto que la deserción para un estudiante es el fracaso para alcanzar una

meta deseada; mientras que para la institución también es un fracaso por no apoyar al estudiante a cumplir con su objetivo de permanencia en la universidad.

Se puede considerar también, que la admisión al sistema educativo superior ecuatoriano es un proceso donde todos tienen iguales oportunidades para participar, pero la selección está basada en el mérito académico de los aspirantes. El marco legal de este proceso se fundamenta en la constitución ecuatoriana vigente, establece que *“El ingreso a las instituciones públicas de educación superior se regulará a través de un sistema de nivelación y admisión establecida en la ley. La gratuidad se vinculará a la responsabilidad académica de los estudiantes”* [19]. En virtud de que en la actualidad los estudiantes no pueden acceder en muchas ocasiones al sistema de educación superior a la carrera deseada, esto dificulta el trabajo de las instituciones de educación superior debido a que los alumnos se retiran en los primeros años por no encontrarse a gusto en una carrera a la que ingresaron sin motivación [20].

Según la UNESCO “Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura”, la deserción estudiantil del nivel de formación universitaria llega a 40%. En Ecuador, en 2014, el 26% optó por abandonar las aulas universitarias [21]. Guadagni [22] en su publicación indica que *“con un sistema de buena selección más el tema de nivelación, la tasa de retención es del 80%. Ahora se hace la selección antes y se evita el desperdicio de recursos”*. Sin embargo, la literatura ha permitido evidenciar que un gran porcentaje de estudiantes en la actualidad no pueden acceder al sistema de educación superior. El sistema de ingreso actual limita la selección del lugar de estudio y la carrera que por vocación se puede acceder, negando el libre ingreso a la educación que es un derecho constitucional al que todo ciudadano puede acceder.

Las causas la deserción estudiantil desde la perspectiva de las universidades son consideradas como efectos negativos que preocupan a las autoridades, por la disminución del presupuesto que podría estar relacionado [23], la institución

universitaria se ve afectada en sus índices de eficiencia; mientras que el estado invierte grandes cantidades de recursos que se pierden cuando un estudiante se retira del sistema de educación pública. La deserción también afecta a las universidades privadas las cuales pierden altas tasas de dinero por la pérdida de estudiantes.

Se ha descubierto varios estudios e investigaciones sobre la detección temprana de posibles estudiantes desertores causada por diferentes problemáticas que originan la deserción estudiantil universitaria en varias etapas. En esta ocasión con la ayuda de la tecnología, modelos matemáticos y técnicas innovadoras de análisis de datos para predecir la deserción estudiantil, se puede establecer causales de deserción estudiantil que afectan a la Universidad Técnica de Cotopaxi.

Formulación del problema

¿Cuáles son los factores que afectan a la deserción estudiantil en la Universidad Técnica de Cotopaxi, desde la perspectiva de expertos?

Objetivos

Objetivo General

- Diseñar un modelo de predicción para determinar factores de deserción estudiantil en la Universidad Técnica de Cotopaxi, aplicando técnicas de minería de datos basada en el proceso Knowledge Discovery in Databases (KDD).

Objetivos Específicos

- Realizar la revisión sistemática de literatura referente a la deserción estudiantil universitaria y técnicas de minería de datos, para la recolección de

criterios sobre deserción estudiantil y minería de datos, a través de fuentes investigación científicas que permita sustentar la pregunta de investigación e hipótesis.

- Diseñar un modelo teórico para determinar factores de deserción estudiantil mediante la perspectiva de expertos y validación mediante proceso experimental a través de regresión lineal simple con mínimos cuadrados ordinarios.
- Validar el modelo de deserción estudiantil mediante el uso de redes neuronales artificiales para determinar el nivel de confianza de los resultados a través del uso de métricas de minería de datos.

Justificación

Para Tinto [12] el ayudar a las instituciones a reducir las tasa de deserción permitirá que la educación superior cambie y tenga una mejor perspectiva hacia la problemática viéndose en la necesidad de construir procesos o sistemas que logren mantener cómodo y seguro a los estudiantes dentro de la institución, puesto que el alumno debe ser considerado como la razón de ser de la universidad.

El incremento de la producción científica identificada, permite establecer que el tema de la deserción en las universidades porte en el campo de la investigación, en donde se han creado muchas líneas de estudio y revisiones en curso, las cuales reflejan la creciente atención al fenómeno que atañe directamente a la educación superior. Esta problemática se ha vuelto más grande provocando que a nivel mundial el número de personas excluidas de una participación activa en la vida educativa sea mayor.

En un estudio del American Institutes for Research [24] contemplo un periodo de cinco años para analizar las afectaciones de la deserción universitaria en el primer año

en las universidades norteamericanas. Los resultados de la investigación arrojan factores de tipo económicos, concluyendo que el estado gastó más de nueve mil millones de dólares para apoyar a los estudiantes que no regresaron a la universidad para cursar su segundo año de estudios. Es importante mencionar que en este análisis no se consideraron la inversión económica familiar o personal de los estudiantes solo fueron considerados fondos estatales que podrían considerarse como excesivos.

Según Donoso, Arias y Donoso [25] mencionan que las instituciones que impulsan políticas importantes para la retención universitaria son caracterizadas como eficientes por estar siempre velando por el éxito de sus estudiantes, además de tener un vasto compromiso con la igualdad, la promoción de los derechos de acceso a la educación y de equidad de procesos. También, señala que la retención no se debe concebir como una política aislada, por el contrario, es una característica de administración eficiente porque conlleva la optimización global de los recursos humanos y materiales. Demostrando así, que la calidad no es un elemento extra o aislado de la educación superior, ya que la calidad es un componente básico de las instituciones sobre el que se rigen procesos tan elementales como la retención y el éxito estudiantiles.

Garzón y Gil [16] menciona que para garantizar la finalización de los estudios superiores en cada una de las instituciones es necesarios establecer políticas y mecanismos efectivos para garantizar satisfactoriamente dicho proceso. Sin embargo, en la mayoría de los casos, no se logrado atacar de manera efectiva el problema, viéndose en la necesidad de desarrollar al unísono mecanismos que identifiquen y atiendan desde el primer momento, de manera preventiva, a los estudiantes que ingresan con ciertos factores de riesgo de abandono de los estudios superiores.

Referente a estudios de deserción estudiantil aplicando minería de datos, hemos encontrado un trabajo realizado por Márquez, Romero y Ventura [26] quienes aborda el problema del fracaso escolar de los estudiantes, en la investigación propone una

metodología para predecir a los estudiantes que se encuentran en riesgo de deserción o de reprobar sus estudios utilizando diferentes técnicas de Minería de Datos.

Según Formia, Lanzarini y Hasperué [27] el uso de las técnicas minerías de datos es muy relevante cuando se tiene información con alta dimensionalidad, relacionados con datos redundantes o irrelevante. Haciendo un buen uso de estas herramientas se puede lograr la selección de subconjuntos de características que sean relevantes y así comenzar a comprender con qué tipo de datos contamos para realizar una investigación.

Por otro lado Jadrić, Željko y Ćukušić [28] analizan a 715 estudiantes que se agrupan en categorías en función de su rango en la inscripción, esta información recolectada es sometida a un análisis riguroso con el software SPSS, se grafican los datos que son seleccionados para describir la estructura de los estudiantes y la tendencia a la deserción. Eckert y Suénaga [29] realizan un trabajo de investigación para identificar factores que influyen sobre la deserción de los estudiantes universitarios mediante algoritmos de clasificación como árboles de decisión, redes bayesianas, donde los autores identifican variables influyentes en la deserción basados en el desempeño académico del estudiante.

El presente proyecto está enfocado en la minería de datos orientado al sector educativo del Ecuador, teniendo como caso de estudio, la Universidad Técnica de Cotopaxi para determinar factores de deserción estudiantil aplicando técnica de minería, a través de un proceso de validación de expertos. El presente proyecto de investigación es factible porque se dispone de los recursos necesarios como son: información relevante de los expertos de la Universidad Técnica de Cotopaxi, herramientas tecnológicas libres para diseñar y evaluar el modelo para la determinación de factores de deserción estudiantil, recursos bibliográficos actualizados y especializados en el tema.

Hipótesis

Si se diseña un modelo teórico para la determinación de factores de deserción estudiantil universitaria bajo la perspectiva de expertos, entonces se podrá obtener una herramienta de apoyo que determine el incremento de las tasas de deserción la cual contribuirá en la toma de decisiones oportunas para las autoridades de la Universidad Técnica de Cotopaxi.

Metodología

Creswell [30] argumenta que la investigación mixta permite integrar, en un mismo estudio, metodologías cuantitativas y cualitativas, con el propósito de que exista mayor comprensión acerca del objeto de estudio. Aspecto que, en el caso de los diseños mixtos, puede ser una fuente de explicación a su surgimiento y al reiterado uso en ciencias que tienen relación directa con los comportamientos sociales. Se retoma a continuación la organización de diseños mixtos planteada por [30] quien señala que es posible organizarlos de la siguiente manera:

- Estrategia secuencial explicatoria: Determina que los resultados cualitativos se utilizan para explicar resultados cuantitativos, el orden es cuantitativo – cualitativo, el énfasis es explicar e interpretar relaciones existentes en los resultados.
- Estrategia secuencial exploratoria: Determina que los resultados cuantitativos se usan para explicar los cualitativos, el orden es cualitativo con análisis, seguido de cuantitativo con análisis.

Por lo que se puede determinar que las aproximaciones metodológicas mixtas permiten, neutralizar o eliminar sesgos de determinados métodos cuando estos son utilizados de forma aislada; se garantiza que los resultados de un método contribuyan

al desarrollo de otros; o que puedan convertirse en una especie de subproceso de otro método, para así proporcionar un mejor análisis a objeto de estudio en nuestro caso aplicando la investigación cuantitativa y cualitativa.

Christ [31] señala que la investigación mediante métodos mixtos se ha fortalecido en los últimos veinte años, y los estudios exploratorios cualitativos, seguidos de estudios confirmatorios, han sido comunes y concurrentes. Al igual que Cameron [32] quien considera que una investigación con métodos mixtos ha ido generando cambios metodológicos entre los investigadores y académicos en una variedad de áreas disciplinarias. Con la complementación metodológica de las investigaciones con un enfoque cualitativo y cuantitativo fueron evolucionando hasta llegar a conceptualizarlas en función de unos nuevos diseños denominados como de modelo y método mixto.

Así, es como los autores antes mencionados señalan que los diseños mixtos permiten, a los investigadores, combinar paradigmas, para optar por mejores oportunidades de acercarse a importantes problemáticas de investigación. En ese sentido, señalan que la investigación mixta se fortaleció, al poder incorporar datos como imágenes, narraciones o verbalizaciones de los actores, que de una u otra manera, ofrecían mayor sentido a los datos numéricos [33]. De la misma forma Pereira [33] afirman que los diseños mixtos permiten la obtención de una mejor evidencia y comprensión de los fenómenos y, por ello, facilitan el fortalecimiento de los conocimientos teóricos y prácticos.

Con este antecedente de proponer una investigación de tipo mixta, la investigación cuantitativa y cualitativa utilizando el instrumento antes mencionado aplicado a experto que están directa o indirectamente relacionados con la deserción estudiantil universitaria, esta información no ayudara a la construcción del modelo conceptual con su respectiva validación.

Por otro lado, investigaciones consideran, que la minería de datos es una disciplina que ha venido tomando fuerza en los últimos años. Las organizaciones han comprendido que los grandes volúmenes de datos que residen en sus sistemas pueden ser analizados y explotados para obtener nuevo conocimiento a partir de los mismos y este sea de gran aporte para la toma de decisiones.

Para Moine, Haedo y Gordillo [34] manifiesta que algunas organizaciones están implementando el proceso de Minería de Datos KDD, mientras que otras han optado por utilizar CRISP-DM puesto que dicha metodología aplica un estándar más específico para este tipo de proyectos. También existen organizaciones que han adquirido productos de la empresa SAS, estas herramientas tiene la particularidad que tiene metodologías SEMMA desarrollada específicamente para este tipo de productos. Otra metodología que cada vez está ganado popularidad es Catalyst (conocida como P3TQ) que debido a su robustez y flexibilidad para adaptarse a distintos escenarios para explorar información.

En los últimos años los esfuerzos en el área de la minería de datos se han centrado en la investigación de técnicas para la explotación de información y extracción de patrones (tales como árboles de decisión, análisis de conglomerados y reglas de asociación). Moine, Haedo y Gordillo [34] menciona que lamentablemente no se ha profundizado en gran escala ejecutar este proceso hasta obtener el “nuevo conocimiento”, es decir, en las metodologías. Las metodologías ayudan a las organizaciones a entender el proceso de descubrimiento de conocimiento basados en una guía para la planificación y ejecución de proyectos de minería de datos de forma sistemática y no trivial.

CAPÍTULO I

FUNDAMENTACIÓN TEÓRICA

1.1 Deserción

Páramo y Correa [4] menciona, que la deserción estudiantil no debe ser conceptualizada solo como el abandono de las aulas de manera definitiva, sino también como el abandono de la formación académica de individuo, independientemente de las condiciones y modalidades en las que estudie, es una decisión personal y no obedece al retiro académico forzoso enmarcado en el bajo rendimiento académico del estudiante, o el retiro por asuntos disciplinares. Según Paredes [35] la descripción de todas las definiciones de deserción estudiantil trata de un fenómeno multicausal, complejo y que aún no se han precisado todas sus las causas. Tinto [18] afirma que la deserción estudiantil universitaria es un fenómeno causal que tiene diferentes factores que confluyen, como aspectos de orden académico, institucional, personal, sociodemográfico.

Gracias a publicaciones realizadas, de artículos que conforman la fuente documental de esta investigación, se han podido reconocer diferentes formas de identificar la deserción estudiantil, obteniendo una amplia lista relacionada con diversas realidades,

(por ejemplo la deserción, la no titulación, la repitencia, el bajo rendimiento académico, la no comprensión de los temas, la exclusión, entre otros) que lo catalogan como una producción social compuesta por dimensiones políticas, sociales, culturales, pedagógicas, personales, cognitivas y psicológicas; enmarcadas en situaciones de índole familiar, material y cultural determinadas por el contexto [3].

1.1.1 Deserción estudiantil universitaria

Para Corena, Lozada y Mahecha [36] manifiestan que la que la deserción estudiantil universitaria se ha convertido en un fenómeno por varias causas y factores atribuida a las instituciones educativas, causando un gran impacto a todo el sistema donde están involucrados tanto los estudiantes, instituciones educativas de nivel superior y el estado, en este sentido, es necesario elaborar estrategias y programas para la prevención y disminución de la tasa de deserción estudiantil universitaria. Según Chinome, Ruiz y Fernández [37] la deserción estudiantil viene siendo un tema muy amplio que se ha tratado de analizar desde las diferentes perspectivas como son: psicología, social, económica y organizacional, por lo que no se ha logrado determinar una definición exacta sobre este fenómeno. En la mayoría de trabajos de investigación coinciden en que la deserción está relacionado con el abandono de los estudios o la prolongación del tiempo para cumplirlos, esta problemática que es causada por variables asociado a diferentes factores relacionados con la deserción estudiantil.

Hablando del Ecuador, en 2016 la Comisión Especial de Estadística de Educación indica que el término más adecuado para utilizarse es “*Deserción estudiantil*” y además ha establecido el cálculo para la deserción estudiantil denominada “*Tasa de deserción estudiantil*” como el número de estudiantes contabilizados al fin de un periodo académico que han abandonado un determinado grado de estudio.

Según Tinto [38] en su obra de estatus paradigmático reconoce la importancia que las características de entrada o preuniversitarias, con las que los estudiantes llegan a las instituciones de educación superior, influyen en la decisión de abandonar la educación de tercer nivel. Estas características incluyen tanto variables socioeconómicas como académicas. La investigación de Vicent Tinto junto con la de Alexander Astin, publicada en 1975 y su teoría de involucramiento [39] marcan un antes y un después en el estudio de la deserción estudiantil. Estas publicaciones son citadas en la mayoría de estudios sobre esta problemática

Según Montes, Almonacid, Gómez, Zuluaga y Tamayo [40] la deserción estudiantil es un fenómeno de gran interés para las Instituciones de nivel superior, entendiéndose como deserción a la interrupción o desvinculación por parte de estudiante del proceso académico de una universidad, al mismo tiempo realiza una investigación conceptual y metodológica de la deserción, analizando la problemática sobre la base de indicadores y características individuales del estudiante como: socioeconómicas, académicas e institucionales.

Existen algunos factores asociados al estudiante que también influyen en la deserción, una de las principales causas tiene que ver con la dificultad del estudiante a integrarse al nuevo ambiente universitario, donde ya no se tiene el acompañamiento que se da en las instituciones secundarias y se pasa a un ambiente de total independencia, que requiere de cierto grado de madurez de estudiante para afrontar ese cambio, por eso “la deserción o la permanencia dependen fuertemente de un exitoso acoplamiento entre la institución y el individuo” [41].

1.1.2 Tipos de deserción

Para Fontalvo, Castillo y Polo [42] existen dos tipos de abandono, los cuales se señalan a continuación:

- *Deserción precoz:* estudiante que habiendo sido admitido en la universidad no se matricula.
- *Deserción temprana:* estudiante que abandona sus estudios en los primeros semestres.
- *Deserción tardía:* estudiante que deja sus estudios en los últimos semestres.

Otro tipo de deserción es respecto al espacio que menciona Fontalvo, Castillo y Polo [42], esta se divide en dos:

- *Deserción institucional:* se refiere al estudiante abandona la universidad.
- *Deserción del programa académico:* es cuando el alumno decide cambiarse a otro programa académico que ofrece la misma universidad.

1.1.3 Factores de deserción

Para iniciar el estudio de fenómeno de la deserción estudiantil propiamente dicha se debe tomar en cuenta los factores que provocan, así empezaremos con un breve análisis, donde se debe considerar que un gran número de desertores que son señalados como causa de un abandono universitario ya sea por situaciones externas o ajenas a su situación. Ya que la responsabilidad propia y la palabra “fracaso”, son muy difíciles de asimilar.

Se debe considerar también, factores externos como pautas principales para podernos adentrarnos en el estudio de casa uno de ellas, y así expandir el criterio del porqué de las deserciones estudiantiles son frecuentes. Cabe mencionar que también existen estudios que señalan que este tipo de factores se han convertido en una excusa general por parte del desertor al no asumir su responsabilidad y acusa el alejamiento de la institución, utilizando a estos factores como chivo expiatorio del sentimiento de culpa que conlleva la decisión ya tomada a la hora de dejar de estudiar [35].

De igual manera, en el Ecuador la deserción de estudiantes de pregrado esta determina por diferentes factores como los de tipo económicos que están relacionados con los ingresos familiares, desempleo de la cabeza de hogar, incompatibilidad entre trabajo y estudio, falta de apoyo familiar, situación económica precaria del estudiante, los factores académicos se relacionan con el bajo rendimiento en hora clase, elección inadecuada de la carrera, repetición de la materia, limitada experiencia académica del colegio, escasa información y orientación a la hora de elegir una carrera. Sin olvidarnos de que también existen factores internos que afecta a cada individuo como por ejemplo: factores de tipo personal como la desmotivación, la insatisfacción académica, problemas de salud, problemas emocionales. Entonces diremos que lo económico, académico, personal y lo vocacional son los puntos claves dentro de los factores más importantes que provocan la deserción estudiantil, por consiguiente se puede establecer que los criterios vertidos por Vicent Tinto donde señala que la deserción tiene estrecha relación con el ámbito social y la integración del estudiante en este.

En nuestra búsqueda de información se logra clasificar como causar externas a los factores económicos, académicos y vocacionales que incluso son analizadas dentro del núcleo familiar, dando lugar a estos como fuentes que exponen al estudiante a expectativas. Lo cual genera que la vida del estudiante en la universidad se torne más compleja, debido a que se encuentra en una presión latente de no defraudar a su familia y así no sentirse fracasado en la búsqueda de una exigencia social, se comienza a dar más valor al punto de vista enfocado en el que dirá la familia en el caso del fracaso estudiantil. También, se logra identificar como causa interna a los factores personales que están relacionados directamente con el comportamiento humano. Paredes [35] realiza un estudio a fondo de cada uno de los factores que se acaban de mencionar, se empezara al igual que otros investigadores hace referencia a la importancia de analizar factores internos seguido de los factores externos, convirtiéndose cada uno de estos puntos como factores fundamentales para la mejor

comprensión acerca de la deserción estudiantil y al mismo tiempo los principales causantes de dicho problema.

Es importante también considerar la posibilidad de establecer análisis de la orientación académica y vocacional, es primordial en la toma de decisión del seguimiento de una carrera, para que el alumno que se va a convertir en estudiante universitario no se desmotive de la carrera que eligió o tenga una decepción en ámbito académico, se considera que para el éxito del individuo debe tener bienestar en el ámbito personal como lo menciona Rojas y González [43]. Una mala condición de salud, insatisfacción académica, inestabilidad emocional o desadaptación social, permite que el estudiante no pueda rendir académicamente de una manera adecuada en la universidad, de tal forma que se fomenta un desánimo lo que produce que se genere la idea de abandonar la carrera escogida en un inicio, es por eso que el ámbito personal es uno de los factores que influye fuertemente a la decisión de desertar.

Dentro de los factores académicos analizados se establece que no es claro el límite entre lo individual y lo académico pues la orientación vocacional, o su ausencia, y el éxito o fracaso en el proceso de adaptación a la vida académica suelen formar parte del mismo transcurso [43]; lo académico se distingue por las calificaciones y los programas ejecutados por parte de los maestros, los cuales influyen en el ánimo del estudiante alentándole a desertar o a seguir motivado por su elección de carrera, pero a lo académico le influye de un modo u otro el factor personal, debido a que si se tiene un problema personal ya sea emocional, de salud, de adaptación o la duda de si se eligió la carrera correcta, va a generar una inestabilidad en el rendimiento académico, incidiendo en el sentimiento de no pertenecer ahí y ejecutar el abandono inmediato de su carrera.

Dentro de los factores socioeconómicos de deserción estudiantil se puede considerar que puede ser uno de los factores más destacados en cuestión a las causas de deserción, debido a que la falta de facilidad de pago y los tantos obstáculos que se

presentan para poder obtener un crédito estudiantil, esto sumado a “la ausencia de estímulos, becas, subsidios, etc. Incrementan el problema socioeconómico” [44], fomentando el abandono del estudiante que se ha visto afectado por la falta de un nivel económico adecuado para solventar su carrera estudiantil, este caso es más frecuente en las universidades privadas, debido a que en ellas para ingresar a la carrera escogida y continuar los diferentes semestres se necesita un porcentaje alto de inversión económica.

Factores Institucionales, factor que incide en un gran porcentaje en la deserción estudiantil universitaria que está relacionada con la baja calidad académica, la falta de programas adecuados para los individuos, la inflexibilidad institucional y académica, la falta de motivación a seguir en la institución, prácticas docentes obsoletas, espacios pedagógico inadecuados, que no aportan de ninguna manera a un buen desarrollo del estudiante [45].

Factores vocacionales, este factor se ve reflejado en la falta de información sobre las carreras académicas de la universidad, no brindando así con la orientación acerca de las exigencias de la carrera a corto y largo plazo, tanto al inicio como en el transcurso de la misma, puesto que el estudiante necesita una constante orientación y apoyo psicológico para la adaptación adecuada [46], influyendo en el estudiante en la toma inadecuada a la hora tomar la decisión para elegir una carrera universitaria, lo cual aumenta la posibilidad de que el estudiante decida retirarse de la misma, porque no fue lo que esperaba.

Recopilando todos y cada uno de los conceptos anteriores se podrá decir que los factores tomados en consideración por los autores, son uno de los puntos clave para poder identificar el fenómeno de la deserción universitaria, este problema viene acarreado un sin número de problemas y consecuencias indeseadas tanto para el estudiante como para las instituciones. Con toda la información recopilada tendremos la oportunidad de conocer más a fondo que la deserción no es un resultado neutro y

tampoco una decisión tomada al azar, más bien es la unión de circunstancias que confabulan para fomentar el progreso de una decisión que en el futuro afectara tanto al estudiante, a su familia, a la sociedad y a la institución donde logro ingresar durante el periodo que determino quedarse.

Revisando la literatura existen varios estudios donde se ha encontrado varios factores asociados a la deserción estudiantil que se presentan a continuación:

- Trayectoria educativa de la institución y acompañamiento al estudiante en su formación.
- Ambientes educativos.
- Ambiente familiar.
- Edad del estudiante.
- Adaptación social del estudiante con sus compañeros.
- Modelos pedagógicos diferentes a los del nivel secundario.
- Horas clases muy rígidas, de alta intensidad temática y tiempos reducidos para dictar dicha temática.
- Evaluaciones extenuantes y exigentes.
- Materias no asociadas ni aplicables al desarrollo profesional del estudiante.
- Factores económicos.
- Cantidad de oferentes (el mercado de la educación).
- La orientación profesional y vocacional de la institución.
- Hijos de padres a los que no les interesa la educación.
- El no aprovechamiento de oportunidades educativas.
- Problemas de disciplina.
- Falta de interés por los estudios.
- Nivel socioeconómico bajo por parte de los padres.
- Baja empatía por el trabajo de sus compañeros.
- Desmotivación hacia la carrera o la Universidad.
- Resistencia a desarrollar actividades curriculares.

- La falta de ganas de adquirir nuevos conocimientos.
- Ausentismo.
- Salud psicossomática.
- Relaciones interpersonales.
- Procedencia de entornos familiares violentos.

Por consiguiente se puede establecer que la deserción estudiantil no es un problema solo del individuo; desde luego que el desertor se convierte en el actor principal, pero no por eso le vamos a declarar como el único culpable de la problemática. Sin embargo se determina que todo converge directamente en el desertor durante todo el proceso de la deserción o abandono de una institución universitaria. Atores de la talla de Tinto, Simpson, Casillas, Crrred, Fallon, Robotham y Hood [47], entre otros, mencionan que los factores asociados con el abandono y/o fracaso académico en la educación superior son múltiples y de diversa naturaleza, por lo que se realiza la siguiente clasificación, tabla 1.

Tabla 1: Variables personales, institucionales y circunstanciales para la retención universitaria.

Variables Personales	Variables Institucionales	Variables Circunstanciales
<i>Variables demográficas</i>	<i>Variables Burocráticas</i>	<i>Interacciones Institucionales</i>
Edad, género, raza, residencia, ingresos familiares, nivel socioeconómico, educación de los padres, expectativas de los padres, apoyo familiar, horas de trabajo y responsabilidades familiares	Misión y política, tamaño de la institución, presupuesto y fondos, reconocimiento institucional y participación	Interacciones burocráticas, Interacciones académicas, Interacciones sociales (actividades extracurriculares, interacción con pares, interacción con el personal de la Facultad)
<i>Variables individuales</i>	<i>Variables Académicas</i>	<i>Interacciones externas a la institución</i>
Estrategias y autorregulaciones académicas (por ejemplo, procrastinación, estilos de aprendizaje); personalidad; motivación al logro; compromiso; expectativas.	Titulación, sistema estructural, sistema normativo, sistema de evaluación de la calidad del profesorado	Circunstancias de la vida Circunstancias laborales Circunstancias familiares y/o socioeconómicas
<i>Experiencias educativas previas</i>	<i>Variables Sociales</i>	
Rendimiento previo, experiencias escolares previas, orientación previa sobre la titulación a cursar.	Sistema social Mecanismos de integración social Becas, apoyos, mecanismos de financiación.	

Fuente: Tomado de Garzón [16].

1.1.4 Modelos de deserción Universitaria

Bajo el enfoque de Braxton [48] y el análisis realizado sobre la deserción y retención pueden ser agrupados en cinco grandes categorías, de acuerdo a las afectaciones que tengan las variables explicativas, ya sean individuales, institucionales o del entorno familiar. De esta manera se reconocen los siguientes enfoques:

- Psicológico
- Económicos.
- Sociológicos
- Organizacionales

Modelo de Vincent Tinto

Así, es como el Modelo de Tinto [38] ha tenido mucho éxito desde 1975 en el que fue publicado, su modelo trata específicamente de deserción universitaria. Este modelo señala que las metas que persigue el estudiante para su educación y sus compromisos institucionales influyen en la deserción. Otros investigadores han propuesto distintos modelos para la retención de los estudiantes enfocados en diferentes perspectivas, pero cada universidad tiene unas características diferentes que deben ser identificadas con el fin de hacer frente al problema. El modelo de Tinto sugiere que los estudiantes al ingresar a la universidad poseen atributos familiares y propios como individuos, pero al momento que el estudiante es admitido por la institución, una serie de factores del propio sistema social lo afectan donde se ven involucrados tanto profesores como también el grupo de compañeros de clase, el modelo es muy eficiente porque está basado en el involucramiento de los estudiantes donde sus metas y el compromiso institucional son fundamentales los mismos que facilitan el entorno social en el que se encuentran [49].

Estos compromisos a su vez serán afectados tanto por factores propios del estudiante y de su entorno, como por experiencias que puedan vivir en la universidad. Todos estos factores determinarán si el alumno decide quedarse en la institución, trasladarse o desertar definitivamente del sistema educativo. De este modo se postula un modelo causal de cinco etapas las cuales observaremos en el gráfico 1.

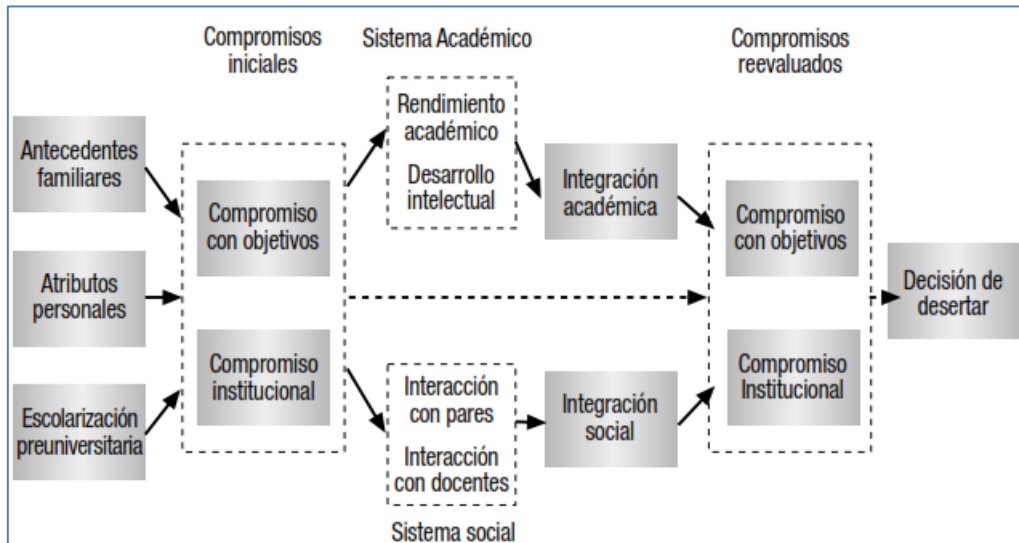


Gráfico 1: Modelo deserción académica de Tinto. Tomado de Himmel [50].

En el modelo propuesto en la gráfica 1, se puede observar los cinco factores que intervienen en la decisión estudiantil, en la primera hace referencia a los atributos previos al ingreso a una institución universitaria como son: los antecedentes familiares, las destrezas y habilidades del individuo y la escolaridad previa. La segunda trata de las metas y compromisos académicos del estudiante. La tercera habla sobre las experiencias institucionales que ha vivido el estudiante relacionados con el sistema académico en el cual se observa como le ha ido en sus estudios y el sistema social el mismo que determina si el individuo se ha adaptado o no al nuevo estilo de vida. La cuarta nos indica la integración del individuo en el entorno universitario, Tinto dice que un estudiante con un rendimiento académico adecuado y que tiene una buena interacción con sus profesores, se sentirá cerca al grupo y no habrá tanta

posibilidad de que piense en desertar. La quinta son las nuevas metas y compromisos que se ha planteado el individuo con relación a todos los factores, si estas están fuertemente adaptadas a su nuevo entorno no habrá duda por parte del estudiante en quedarse o no en la carrera, es así que la suma de todas estas variables, determinara el futuro del estudiante.

Modelo de Fishbein y Ajzen

Los enfoques psicológicos de la retención estudiantil se centran en rasgos de personalidad que diferencian a los alumnos que completan los estudios que aquellos que no lo completan. Así, es como Fishbein y Ajzen [51] señala que la decisión de desertar o mantenerse en un programa de estudio se ve influida por las conductas previas, llevando a la conformación de una intención conductual, la que finalmente se traduce en comportamiento de individuo. Esta concepción puede explicar la deserción voluntaria, como la retención estudiantil, como se muestra en el gráfico 2.

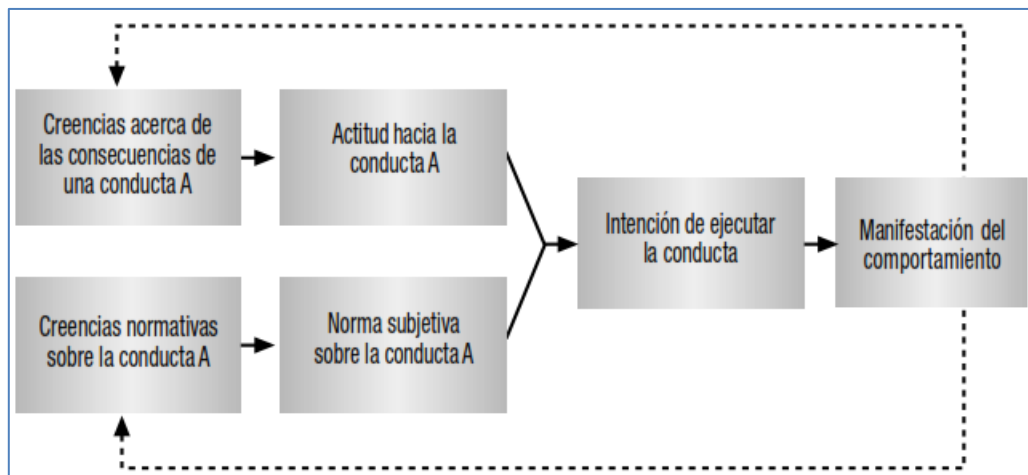


Gráfico 2: Modelo de Fishbein y Ajzen. Tomado de Himmel [50].

Modelo de Ethington

Por otro lado Ethington [52] construyó un modelo más completo, basándose en las anteriores e incorporando una teoría más general sobre las “conductas de logro” de Eccles [53], una de las primicias de Eccles es que el rendimiento académico influye en desempeños futuros, como en sus metas, valores y expectativas de éxito. Ethington [52], muestra en el gráfico 3, que encontró empíricamente que el nivel de aspiraciones tenía un efecto directo sobre los valores, además pudo observar que las expectativas de éxito estaban implicadas por el autoconcepto académico y la percepción de las dificultades académicas. Tanto los valores como la expectativa de éxito influyen a su vez en la permanencia en la universidad.

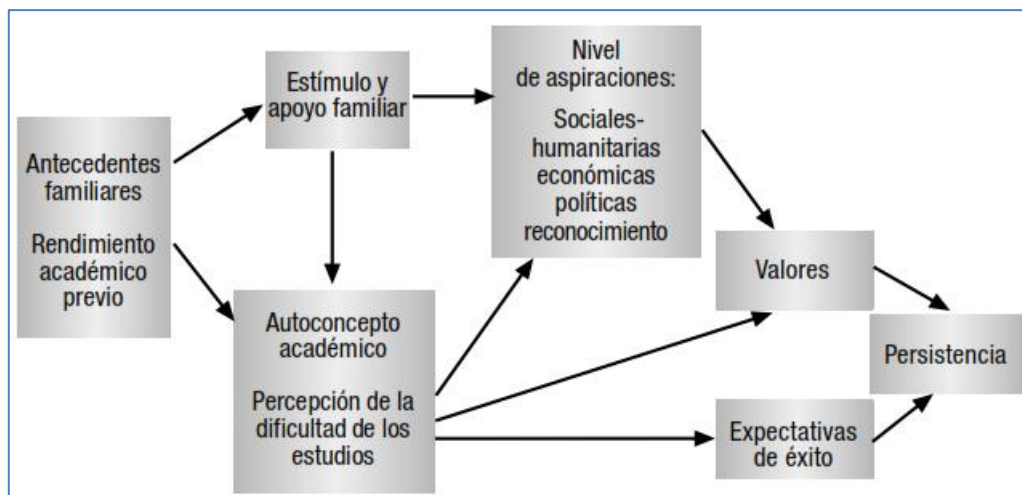


Gráfico 3: Modelo de Ethington. Tomado de Himmel [50].

1.2 Proceso de extracción de conocimientos

Para Matos, Chalmeta y Coltell [54] el conocimiento es una combinación de las experiencias, información contextualizada, que integrando a un proceso que permita evaluar e incorporar nuevas experiencias e información que transformase en conocimiento. La mayoría de Organizaciones definen al conocimiento como aquella

información que permite generar acciones asociadas para satisfacer las demandas del mercado y estar un paso adelante de la competencia.

Según Valcárcel [55] La extracción de conocimiento de un gran volumen de datos, está implícita, puesto que se trata de interpretar grandes cantidades de datos para encontrar patrones o relaciones entre sí, haciendo uso de diferentes técnicas de aprendizaje y la estadística, aplicando a un conjunto de datos detenido para su evaluación. El objetivo es incorporar el resultado “Conocimiento” obtenido en algún sistema para la toma de decisiones.

1.2.1 Minería de datos

Para Britos y Garcia [56] determina que la minería de datos es entendida como el proceso para descubrir conocimiento valioso relacionado entre sí, basándose en las asociaciones, cambios, anomalías y estructuras significativas a partir de grandes volúmenes de datos almacenados en archivos, bases de datos, o cualquier otro medio de almacenamiento de información digital. Para un eficiente proceso de minería de datos requiere de actividades previas destinadas a preparar los datos de manera homogénea y al mismo tiempo la aplicación de algoritmos para desarrollar dicha tarea. Esta primera etapa es también conocida como ETL (Extract, Transform and Load) [58]. Un proceso completo de aplicación de técnicas de minería, mejor conocido como proceso de descubrimiento del conocimiento de una gran variedad y tipos de datos establece a la minería de datos como una etapa del mismo proceso [59].

De igual manera el proceso de extracción de conocimiento a partir de un conjunto de datos, diversos autores lo han definido de diferentes maneras o con diferentes palabras, pero todos refieren las mismas ideas. Al citar una de las definiciones; la minería de datos es proceso de descubrir conocimientos interesantes como: patrones, cambios, anomalías, asociaciones y estructuras significativas a partir de grandes cantidades de datos almacenadas en base de datos, o cualquier otro medio de

almacenamiento de información [56]. De igual forma el proceso Knowledge Discovery in Databases (KDD) [57] es utilizado para el descubrimiento del conocimiento en base de datos, que se encarga de la preparación de los datos que ayuda a la interpretación de los resultados obtenidos, dando así sentido a cada uno de los patrones encontrados.

1.2.2 Metodología (CRISP-DM) para proyectos de minería de datos

Cross Industry Standard Process for Data Mining (CRISP-DM), es la guía de referencia más ampliamente utilizada en el desarrollo de proyectos de Data Mining, como se puede constatar en la gráfica presentada en el gráfico 4. Esta gráfica, publicada el año 2007 por kdnuggets.com [60], representa el resultado obtenido en sucesivas encuestas efectuadas durante los últimos años. Una revisión y crítica de los modelos de minería de datos en 2009 llamó a CRISP-DM el “estándar de facto para el desarrollo de la minería de datos y los proyectos de descubrimiento de conocimiento” [61].

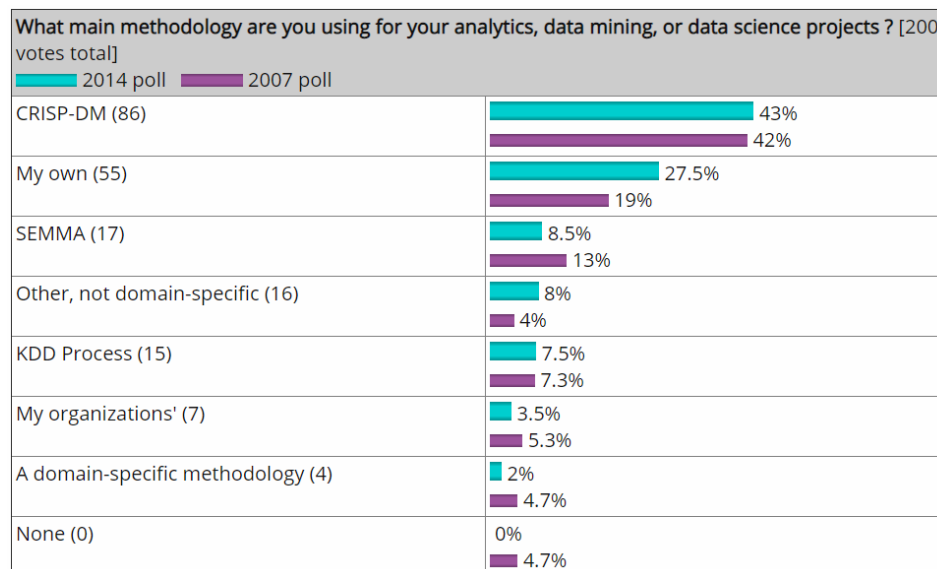


Gráfico 4: Encuesta de uso de metodologías en proyectos de minería de datos. Tomado de [60].

CRISP-DM, es la metodología más recomendada debido a que su fase inicial es el reconocimiento y entendimiento del negocio y de sus datos; y finaliza el despliegue de los modelos que describen el comportamiento de las variables del negocio entorno al problema formulado.

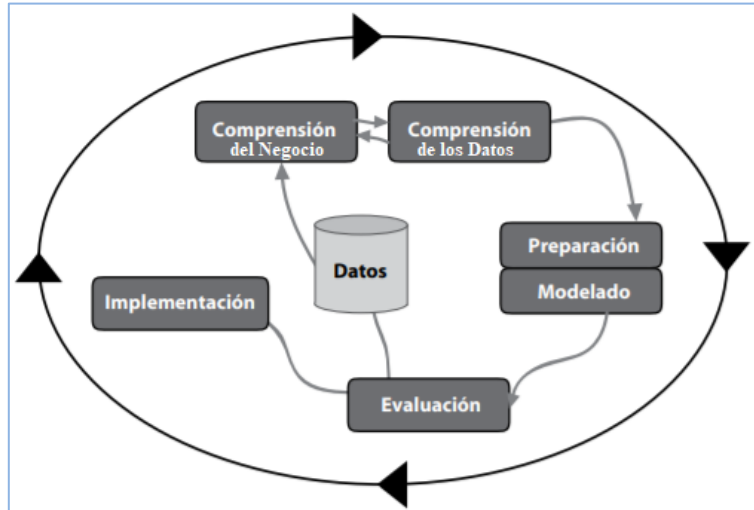


Gráfico 5: Modelo CRISP-DM. Tomado de [62].

A continuación se describen las 6 fases en que se divide CRISP-DM basado en el autor [62].

- **Definición de necesidades del cliente.-** Esta fase inicial es muy importante, ya que se enfoca en la comprensión de los objetivos de proyecto. Primero hay que diseñar un plan preliminar para alcanzar los objetivos planteados en el problema, los mismos que serán resueltos aplicando minería de datos para que estos se conviertan en conocimiento.
- **Estudio y comprensión de los datos.-** En esta fase es importante la comprensión de datos, comienza con la recolección de datos iniciales y continúa al otro nivel donde permiten familiarizarse con los datos donde se identifican los problemas de calidad de estos, también se descubre el conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis para ser resueltas en cuanto a la información oculta que existe.

- **Análisis de los datos y selección de características.**-En esta etapa, básicamente se usa para adaptarnos a la técnica de minería de datos, mediante la visualización de los datos y la búsqueda de relaciones entre las variables para construir el conjunto final de datos. Estos datos se utilizarán en las herramientas de modelado, Las tareas de esta fase incluyen la selección de registro, tablas y atributos, así como la limpieza y transformación de los datos que estén listos para ser introducidos a las herramientas que modelan.
- **Modelado.**- En esta fase se procede a la selección de técnicas de modelado más apropiadas para el proyecto de Minería de Datos. Las técnicas a utilizar en esta fase se eligen en función de los siguientes criterios: disponer de datos adecuados, ser apropiada para el problema, cumplir los requisitos del problema, el conocimiento de la técnica y el tiempo adecuado para obtener un modelo estable. Si el modelo construido no resuelve el problema planteado, casi siempre se termina volviendo a la fase de preparación de datos, para comenzar de nuevo.
- **Evaluación.**- Para llegar a esta etapa del proyecto, se debió haber construido uno o varios modelos que alcanzaron calidad suficiente desde la una perspectiva de análisis de datos. En esta fase es muy importante evaluar y revisar a fondo los resultados obtenidos, comparar el modelo obtenido con cada uno de los objetivos del negocio. Estos pasos son muy importantes antes de poner en producción al modelo.
- **Despliegue.**- Aquí es donde el conocimiento obtenido se transforma en acciones dentro del proceso de negocio, ya sea observando el modelo y resultados, o aplicándolo a múltiples grupos de datos o como parte del proceso, en esta la última fase de la metodología es donde el conocimiento obtenido en todo el proceso tendrá que organizarse y presentarse al cliente para que pueda hacer usos del mismo.

La metodología es muy fácil de comprender, puesto que existe mucha información sustentada en artículos, también se puede comentar que para determinar el éxito o el fracaso en este tipo de proyectos mucho depende del cómo se ejecute el ciclo de vida del mismo.

1.2.3 Proceso de minería de datos (KDD)

Para Riquelme, Ruiz y Gilbe [63] en el área de las bases de datos se han venido realizando varios trabajos de investigación relacionados con el Knowledge Discovery in Databases (KDD), también comentan que aplicando adecuadamente este proceso conseguimos la extracción de conocimiento a partir de datos dispersos, marcando un cambio de paradigma siendo lo más importante el conocimiento útil que seamos capaces de descubrir a partir de un determinado conjunto de datos. Giraldo y Jiménez [64] manifiesta que KDD es un proceso que analiza rápidamente grandes cantidades de datos que sobrepasa la capacidad humana a la hora de transformar los datos con precisión, al mismo tiempo se encarga de la preparación de la información y de la interpretación de los resultados obtenidos, siendo este conocimiento de gran utilidad dentro de una problemática ya determinada.

Para Santamaria [65] el proceso de KDD abarca varias etapas en su realización, desde la selección de datos que pueden ser necesarios para descubrir conocimiento, hasta visualizar los resultados de dicho descubrimiento. El principal proceso dentro del KDD es la minería de datos ‘Data Mining’, que es la responsable de buscar, descubrir y extraer el conocimiento desde los datos.

Recolectado las opiniones de los autores tenemos que término ‘KDD’ es identificado como un proceso empleado para el descubrimiento y extracción de conocimiento nuevo, no obvio a partir de un conjunto de datos grande y variado, el cual está

conformado por relaciones y patrones entre los elementos que conforman dichos datos.

El proceso KDD en forma general se ve reflejado en la gráfica 6.

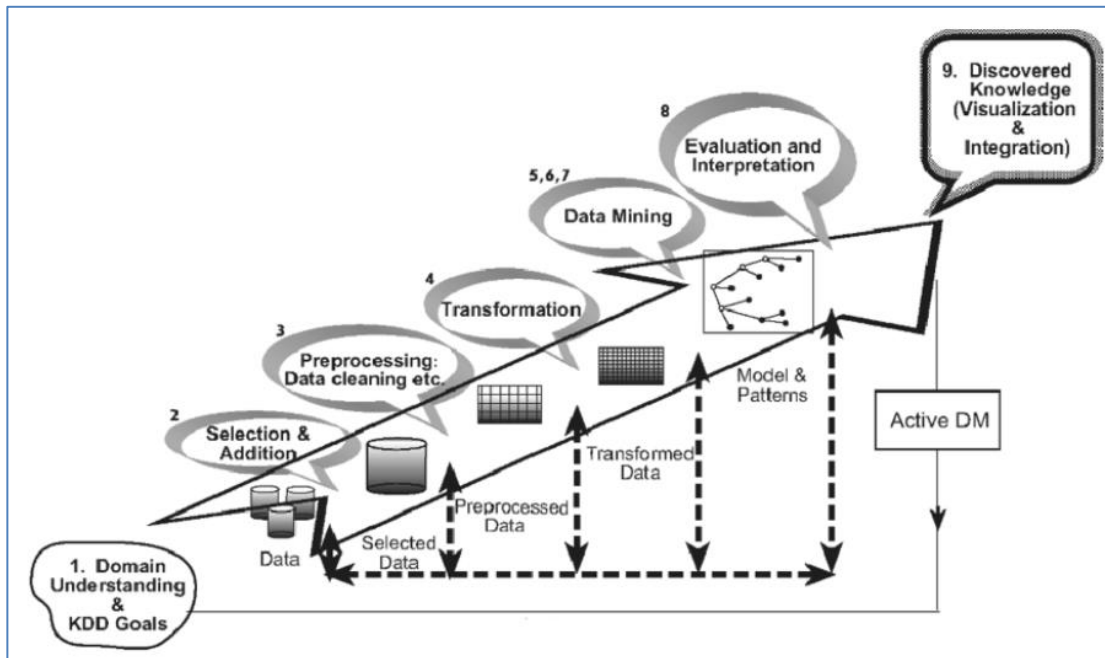


Gráfico 6: El proceso de descubrimiento de conocimientos en bases de datos (KDD). Tomado de Maimon y Rokac [66].

1.2.4 Fases de la minería de datos (KDD)

Según Nigro, Xodo, Corti y Terren [57], las etapas del proceso KDD se dividen en 7 fases que son:

- **Entender el dominio de aplicación**, cuál es el problema a resolver, y cuáles son los objetivos.
- **Seleccionar del conjunto de datos**, en esta etapa se realiza la limpieza y preparación de los datos extraídos desde las distintas fuentes de datos. Para su

mejor compleción, datos más manejables que serán útiles para las fases posteriores.

- **Implantación** del almacén de datos que permita la búsqueda y visualización previa de sus datos, para comprender mejor los aspectos que puede interesar y ser estudiados a profundidad.
- **Selección, limpieza y transformación de los datos**, en esta fase permite la transformación, tratamiento preliminar de los datos, y generación de nuevas variables a partir de las ya existentes. En este paso también se realizan operaciones de normalización de datos o agregación, consolidando así los datos para la siguiente fase.
- **Seleccionar y aplicar el método y algoritmo de minería de datos apropiado**. Elegir las herramientas de data mining adecuadas al problema a resolver, teniendo en cuenta el objetivo (predecir, explicar, clasificar, agrupar, etc.).
- **Interpretación y Evaluación**. En esta fase se identifican las variables, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos como son: Evaluación, transformación, interpretación y representación, de los patrones obtenidos y que son realmente interesantes.
- **Difusión** y uso del nuevo conocimiento.

1.2.5 Modelos para la minería de datos

Los modelos generados para el análisis de datos pueden ser de dos tipos:

- **Modelos de Análisis Descriptivos:** El principal objetivo de este tipo de modelamiento es hallar patrones interpretables que describan los datos. Estos pueden ser reglas asociativas, agrupaciones de datos, similitudes entre los datos; se puede encontrar además conjuntos de segmentos.
- **Modelos de Análisis Predictivos:** Este modelamiento utiliza algoritmos estadísticos y de análisis para predecir valores desconocidos o futuros de unas variables basándose en datos existentes de otras variables.

1.2.6 Técnicas para la minería de datos

Para la minería de datos, las técnicas pueden corresponder a dos tipos; de aprendizaje supervisado y aprendizaje no supervisado; las técnicas de aprendizaje supervisado utilizan datos previos o de entrenamiento con el fin de autoajustar sus funciones hasta obtener un resultado equivalente o aceptable a la realidad [67] por otro lado las técnicas de aprendizaje no supervisado por lo que tratan a los datos ingresados como un conjunto de variables aleatorias sobre los cuales puede encontrar patrones [68]. Según [69] estas son las técnicas comúnmente utilizadas:

1.2.6.1 Aprendizaje supervisado

- **Clasificación.-** es una de las aplicaciones más comunes para la minería de datos. Corresponde a una tarea que se repite con frecuencia en la vida cotidiana. Por ejemplo podemos desear clasificar un proyecto estudiantil en las siguientes calidades: distinción, mérito, pasar o fallar. En todos los casos mencionados la clasificación se utiliza con el fin de predecir una etiqueta.
- **Predicción numérica o Regresión.-** la regresión es utilizada con la finalidad de realizar una predicción numérica; tal como la predicción de ventas de una compañía o el comportamiento de las acciones de una empresa. Algunas

formas de hacerlo es utilizando (Regresión lineal, Series temporales, **Redes neuronales artificiales**).

1.2.6.2 Aprendizaje No supervisado

- **Reglas de asociación.**- una forma de aplicación de esta técnica es el “análisis de canastas de mercado”, donde se busca obtener los productos relacionados que los consumidores compran en un negocio a través del tiempo.
- **Clustering.**- los algoritmos de clúster examinan los datos en búsqueda de grupos de ítems que son similares; estas similitudes son medidas por medio de distancias euclidianas si los atributos son continuos, pero pueden utilizar otros tipos de medidas específicas para problemas específicos. Como características cabe señalar que los objetos dentro de un clúster son similares entre sí; y los objetos en clústeres separados son menos similares entre sí. Un ejemplo de este algoritmo es el K-Medias [70].

1.2.7 Redes neuronales

Para Basogain [71] las redes neuronales son técnicas analíticas que permiten modelar el proceso de aprendizaje de una forma similar al funcionamiento del cerebro humano a nivel conceptual, básicamente, la capacidad de aprender a partir de nuevas experiencias. Estas técnicas tuvieron un desarrollo impresionante en la última década. Camana [72] sostiene que el comportamiento de una red neuronal es de forma muy parecido al cerebro humano aprendiendo de las experiencias y del pasado, son capaces de detectar, aprender patrones y características, una vez que una red neuronal es entrenada adecuadamente puede realizar predicciones, clasificaciones y segmentación, es muy potente al adaptarse y evolucionar ante diferentes realidades.

Una de las ventajas de trabajar con redes neuronales para construir modelos de predicción es que no necesitan trabajar con datos completos, incluso se puede trabajar con datos paradójicos, esto puede resultar un problema o una ventaja, todo dependerá del problema a resolverse. Además, estas técnicas poseen dos formas de aprendizaje: supervisado y no supervisado.

1.2.8 Tipos de redes neuronales

Perceptrón.- Este tipo de red consiste en sumar las señales de entrada y multiplicar por los valores de pesos escogidos aleatoriamente, este valor es comparado con un patrón para determinar si la neurona se activa o no, si el valor comprado es mayor toma el valor de salida 1 caso contrario toma el valor de 0 [73].

Perceptrón Multicapa.- Este tipo de red se caracteriza por utilizar aprendizaje supervisado y es la más utilizada en diferentes aplicaciones, la **MLP (MultiLayer Perceptrón)**, está formado por una capa de entrada, una o varias capas ocultas y una capa de salida. Este tipo de red neuronal usa para su entrenamiento el algoritmo de propagación hacia atrás, conocido también como retropropagación del error o regla delta generalizada [74].

- **Backpropagación.-** Es un algoritmo de entrenamiento que se encarga de corregir el valor de los pesos entre las conexiones de las neuronas mediante la retropropagación del error desde la capa de salida hacia atrás (capa de entrada) a través de las capas ocultas, el número de capas ocultas a utilizarse dependerá del problema que se vaya a resolver.

Máquina de Boltzmann.- Este tipo de redes son capaces de representar y resolver problema complicados de combinatoria y son muy útiles para el reconocimiento de patrones, pues tratan de completar partes que no se conocen [75].

Máquina de Cauchy.- La máquina de Cauchy es una versión mejorada de la máquina de Boltzmann, pues considera funciones alternativas de probabilidad y ajuste de temperatura, aunque las dos presentan una arquitectura y funcionamiento similar. La principal ventaja de la máquina de Cauchy radica en su rapidez de convergencia [76].

1.2.9 Deep Learning

Mohsen, El-Dahshan, El-Horbaty y Salem [77] manifiesta que el Deep Learning al ser una poderosa herramienta de aprendizaje profundo ha ido ganando mucho interés en los últimos años para resolver problemas complejos, así es como la red neurona profunda se viene utilizando en distintos tipos de aplicaciones en el mundo real, procesando grandes volúmenes de datos para el descubrimiento del conocimiento, para realizar predicciones a partir de los mismos.

Para Nielsen [78] Deep Learning o aprendizaje profundo es un poderoso conjunto de técnicas para el aprendizaje basado en redes neuronales, las mismas que están compuestas de un número determinado de neuronas en niveles jerárquicos. En una primera instancia o en la primera jerarquía la red aprende algo simple y luego pasa al siguiente nivel. En el segundo nivel la red toma la información sencilla analizada en el primer nivel y la combina con una información más compleja y así pasa a un tercer nivel, este proceso se hace de manera recurrente hasta obtener la salida deseada.

Según Schmidhuber [79] manifiesta de que el aprendizaje profundo es parte de las redes neuronales, así determina que dependiendo del problema y como están conectadas las neuronas, dicho problema o comportamiento puede requerir largas cadenas de conexiones por cada nivel, esto tiene como resultado una gran precisión en el aprendizaje de la red.

1.3 Conclusiones del Capítulo I

- En el presente capítulo se realiza la búsqueda y el análisis de fuentes de literatura científica relacionados con la deserción escolar universitaria, factores de deserción, modelos de deserción, llegando a la conclusión que la deserción estudiantil es el abandono de una institución de educación superior ya sea pública o privada, la cual puede suscitarse por voluntad propia del alumno o por que sintió obligado por múltiples factores.
- Con base en fuentes de información que se revisó para la presente investigación, también podemos señalar que la permanencia de un estudiante en la universidad dependerá mucho del individuo si logra consigue el equilibrio entre los factores académicos, motivacionales e interrelacionales con la institución, si uno de estos factores se ve afectado, impide que el estudiante permanencia en la institución, provocando así la deserción universitaria sea más recurrente.
- Con la revisión de la literatura científica se ha logrado identificar factores de deserción universitaria en sus diferentes dimensiones como son: individuales, académicos, institucionales y socioeconómicos, los mismos que son analizados de forma rigurosa por los diferentes autores.

CAPÍTULO II.

DETERMINACIÓN DE FACTORES DE DESERCIÓN DESDE LA PERSPECTIVA DE LOS EXPERTOS.

2.1 Recopilación de información

Para Rodríguez [80] en toda investigación es necesario llevar a cabo el proceso de recolección de datos e información, de manera que este se transforma en un paso fundamental para tener éxito en el proceso de investigación. El llevar adecuadamente la recolección de información y la selección del método de recolección de datos es una tarea que todo investigador debe conocer y debe tener mucha práctica con ella.

La recolección de datos es de suma importancia, puesto que se considera como la medición, es una precondition para obtener el conociendo científico de los datos, ya que estos expresan una abstracción del mundo real.

Según Hernández [81] existen instrumentos para la recolección de datos con los cuales el investigador se vale para acercarse a los problemas y extraer de ellos información. El instrumento sintetiza en sí toda la labor previa del investigador, sirve para seleccionar datos que corresponden a los indicadores y por lo tanto a las variables y conceptos utilizados en nuestra investigación.

De esta manera, se considera que la recolección de datos es un paso fundamental, ya que existen algunas equivocaciones o mala práctica en el proceso de desarrollo de investigaciones y que pueden influir en su resultado y pueden ser no objetivas. En toda investigación este paso es uno de los más fundamentales, ya que al momento de realizar la recolección de datos el investigador debe tener ya un bosquejo de que espera obtener con su investigación. Analizando el objetivo de la presente investigación se ha llegado a la conclusión de utilizar la encuesta como el instrumento para la recolección de información y el desarrollo de la investigación

2.1.1 La Encuesta

Para Cohen y Gómez [82] la encuesta es una representación en la que se emplea la entrevista como herramienta para su estructuración. Es un muestreo en el que la población responde interrogantes y se añaden a una categoría en específico con el fin de arrojar un porcentaje total y varios segmentos en los que se denota cada uno de los focos de interés de la encuesta. Por consiguiente se utiliza esta valiosa y más representativa técnica dentro de la investigación, aprovechando sus grandes bondades como son: la alta representatividad, bajos costos, recopilación de datos convenientes, la buena significancia estadística, resultados precisos. La encuesta es realizada para obtener información primaria, a partir de un número representativo de actores de una población, para lograr proyectar sus resultados sobre la totalidad de los actores involucrados [81].

La encuesta fue diseñada para obtener la perspectiva de expertos que se encuentra relacionadas directamente con la problemática de la deserción estudiantil en la Universidad Técnica de Cotopaxi, para poder evaluar los puntos fuertes o débiles de nuestra variable dependiente que es la deserción universitaria, la cual consta de tres secciones que en su conjunto contienen 37 preguntas, como se detalla a continuación en la tabla 2.

Tabla 2: Secciones y preguntas del cuestionario.

Secciones del Instrumento	Nº de Preguntas (máximo)
Sección 1: Caracterización de la Institución y de personas que llena el cuestionario.	11
Sección 2: Compuesta de preguntas que determina la influencia de factores que inciden en la deserción estudiantil.	12
Sección 3: Preguntas que complementan el trabajo investigación.	14
Total Preguntas	37

Elaborado por: Investigador

La encuesta estuvo dirigida a personas con conocimientos y experiencia en temas de deserción estudiantil, los cuales han palpado desde el lado humano y profesional la deserción estudiantil universitaria, como son los tutores, docentes del área pedagógica, psicólogos, personal de bienestar universitario, director o coordinador de carrera, decano o sub decano que son representados, los resultados de este proceso se presentan en la tabla 3.

Tabla 3: Población objetiva, para el diseño muestral.

Diseño Muestral		
Población Objetiva	Detalle Población Objetiva	Descripción Requerimientos
326 Expertos de la Universidad Técnica de Cotopaxi N = 326 expertos	Actores de la Universidad: <ul style="list-style-type: none"> · Tutor · Director o Coordinador de Carrera · Miembro del Equipo de Gestión de la Calidad Institucional o de Carrera · Decano o Subdecano de Facultad · Director de Bienestar Universitario · Docente en el área de Educación · Psicólogo(a) 	<ul style="list-style-type: none"> · Encuesta Impresa · Diseño de encuesta Web para (Ingreso de respuestas)

Elaborado por: Investigador

En el mes de noviembre del año 2019 se realizó la toma de muestra de la investigación sobre los factores que influyen en la deserción estudiantil en la Universidad Técnica de Cotopaxi. En esta sección se describe el proceso de la

encuesta realizada, metodología de contacto con los encuestados, herramienta utilizada para el almacenamiento de los datos, procedimiento para construir el dataset.

2.1.2 Contacto con los encuestados.

Con el objetivo de tener un contacto directo con los expertos, se elabora un proceso a seguir por parte de los investigadores, el mismo que se cumplió en su cabalidad, ya que para todas las encuestas realizadas se mantuvo el contacto directo con cada uno de los expertos de los dos Campus, Matriz y Salache de la Universidad Técnica de Cotopaxi, como se muestra en el gráfico 7. Considerando el objeto y objetivo de este estudio, se procedió a determinar y caracterizar a la población a ser encuestada. Se puede decir que en la Universidad existe un gran interés por la problemática a investigar, por lo que la población estuvo dispuesta a dar todo el apoyo para obtener la información, dando un número total de encuestados de 326 personas.

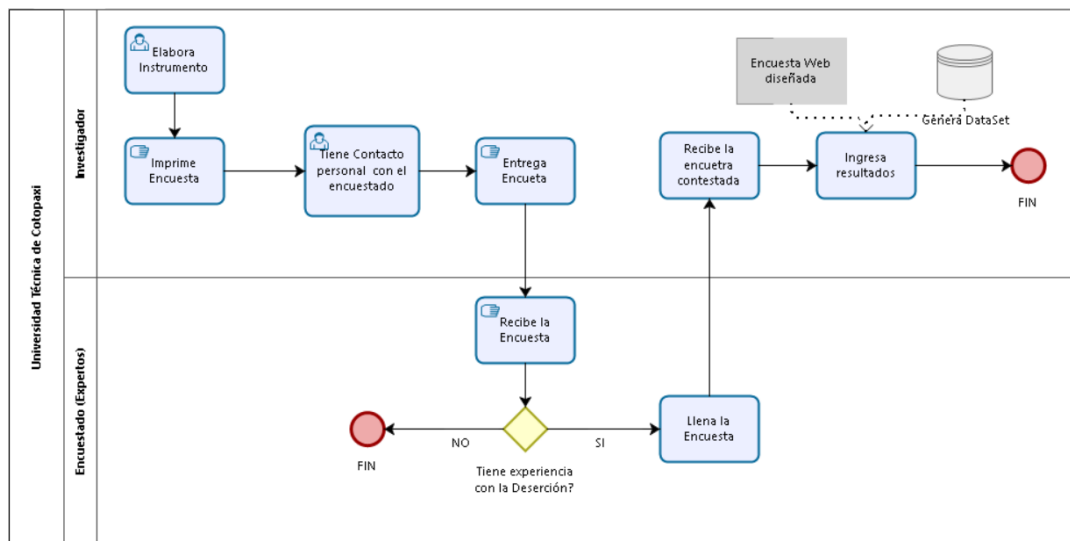


Gráfico 7: Proceso de contacto con los encuestados.
Elaborado por: Investigador

2.1.3 Ingreso de resultados de la encuesta a la plataforma Web

Con la ayuda de las herramientas google-docs y blogspot se implementó el cuestionario con las 37 preguntas en formato electrónico con la finalidad de hacer uso del internet, para acceder vía web al cuestionario y realizar el ingreso de las encuestas por cada uno de los expertos y así retroalimentar nuestro dataset que será analizado durante todo el proceso de la investigación. Los gráficos 8 y 9 presentan el proceso de recolección de datos a través de la encuesta.

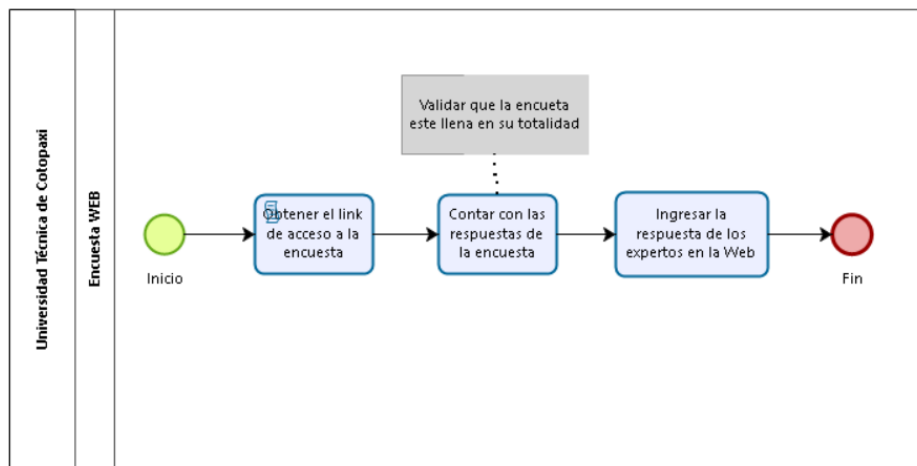


Gráfico 8: Diagrama de la aplicación web desarrollada para ingresar los resultados- Elaborado por: Investigador



Gráfico 9: Encuesta publicada en la web. Elaborado por: Investigador

Todo el cuestionario independiente del perfil del experto, está dividido en tres secciones: Sección 1 está relacionada con la caracterización de la institución y de las personas que llenan el cuestionario. Sección 2 corresponde a preguntas que determinan la influencia de factores que inciden en la deserción estudiantil, y por último la sección 3 contiene preguntas que complementan el trabajo de investigación. El instrumento diseñado nos ayuda a validar el ingreso de las respuestas, puesto que en cada sección las respuestas deben ser respondidas obligatoriamente en forma secuencial. Cada sección valida que se conteste todas las preguntas para seguir a la siguiente sección del cuestionario, caso contrario se señalara la pregunta y se pintara de color rojo indicando la obligatoriedad de responder dicha pregunta. Con se asegura que todas las respuestas sean ingresadas al cuestionario web.

Tasa de respuestas.- Para el estudio de la determinación de factores de deserción estudiantil en la Universidad Técnica de Cotopaxi utilizando minería de datos, fue posible recabar la opinión de **326 expertos**.

2.1.4 Tabulación de la información almacenada.

Con la ayuda de la estadística descriptiva sé realizar la tabulación de cada pregunta de la primera sección de la encuesta realizada. Para procesar los datos se utilizará las siguientes herramientas:

- Google-docs
- Software R
- Software SPSS
- Software Weka

Con la ayuda de la herramienta SPSS se procede a realizar el análisis de confiabilidad del instrumento de investigación utilizando en el presente trabajo. Como se puede observar en la tabla 4, que el valor del Alfa de Cronbach es de $0.925 > 0.8$ el mínimo

aceptable, con los resultados obtenidos se puede considerar que el instrumento utilizado es confiable y válido para la presente investigación.

Tabla 4: Prueba de confiabilidad del instrumento de investigación

Resumen del procesamiento de las observaciones			
		Nº	Percentage %
Casos	Válidos	326	100,0
	Excluidos	0	,0
	Total	326	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad	
Alfa de Cronbach	Nº de elementos
0,925	89

Elaborado por: Investigador

2.1.4.1 Análisis descriptivo de la población estudiada

Para aplicar el análisis descriptivo a la primera sección del instrumento de investigación donde se encuentran las variables como: sexo, edad, años de experiencia en deserción estudiantil, campus, función que desempeña en la institución, profesión del experto, entre otros, se utiliza Rstudio y Google-docs, Ver Anexo 1.

2.1.5 Confiabilidad de los datos

Según [81] la confiabilidad de los datos se da cuando un instrumento se aplica repetidas veces al mismo sujeto u objeto de investigación, por lo cual, se deben obtener resultados iguales o parecidos dentro de un rango razonable, es decir, que en la muestra no se perciba distorsiones. En la segunda y tercera sección del instrumento aplicado se recolectó información valiosa, la misma que ha sido transformada en un dataset con el propósito de facilitar el procesamiento de la información recopilada.

Para determinar la confiabilidad de los datos recolectados de los expertos se utilizó la herramienta Rstudio con el lenguaje de programación R, se procede a aplicar el coeficiente estadístico de Alfa de Cronbach para determinar la fiabilidad de los datos que se utilizaran en la construcción de nuestro modelo de deserción estudiantil. Con las bondades que tiene la herramienta Rstudio para realizar el proceso de validación de los datos, procedemos a exportarlos con la herramienta seleccionada, utilizando la función **alpha()**, como se muestra en los gráficos 10 y 11.

```

#####
#Programa para análisis de fiabilidad de las variables padres##
#Utilizando Alfa de Cronbach
#Edwin G. Chimba
#####
#instalación e paquetes
install.packages("xlsx")
library(xlsx)
install.packages("psych")
library(psych)
#Leer datos y asignar a la variable VariablesPadres
VariablesPadres = read.xlsx("AlfaP.xlsx",1)
View(VariablesPadres)
#Alfa de Cronbach de las variables Padres
alfaPadres <- alpha(VariablesPadres)
#Leer análisis
alfaPadres

```

```

> alfaPadres
Reliability analysis
Call: alpha(x = VariablesPadres)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
0.92 0.93 1 0.13 13 0.0059 3.5 0.39 0.12

lower alpha upper 95% confidence boundaries
0.91 0.92 0.94

```

Gráfico 10: Alfa de Cronbach variables padres.
Elaborado por: Investigador.

```

#####
#Programa para análisis de fiabilidad de las variables hijos##
#Utilizando Alfa de Cronbach
#Edwin G. Chimba
#####
#instalación e paquetes
install.packages("xlsx")
library(xlsx)
install.packages("psych")
library(psych)
#Leer datos y asignar a la variable VariablesPadres
VariablesHijos = read.xlsx("VariablesHijos.xlsx",1)
View(VariablesHijos)
#Alfa de Cronbach de las variables Padres
alfaHijos <- alpha(VariablesHijos)
#Leer análisis
alfaHijos

```

```

> alfaHijos
Reliability analysis
Call: alpha(x = VariablesHijos)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
0.91 0.91 0.98 0.13 10 0.0073 0.7 0.15 0.12

lower alpha upper 95% confidence boundaries
0.89 0.91 0.92

```

Gráfico 11: Alfa de Cronbach variables hijos.
Elaborado por: Investigador.

Como se puede observar en el gráfico 10, el grupo de datos que hacen relación a las variables padres el raw_alpha (coeficiente de Alfa con las puntuaciones observadas) es de 0.92 este valor es usado para evaluar la confiabilidad de los datos. Así también tenemos que para el grupo de datos relacionados con las variables hijos gráfico 11 el

raw_alpha corresponde a 0.91. Sabiendo que los valores arrojados por la función alpha() cuando más cercanos a 1 determinan la consistencia de los datos, para fines prácticos se determina que los valores del Alpha iguales o mayores a 0.7 se consideran aceptables, mayores a 0.8 son buenos y mayores a 0.9 son excelentes. Determinado que los dos conjuntos de datos (padres – hijos) almacenados cada dataset, para la presente investigación son altamente confiables, pues sus valores son mayores al 0.8, el mínimo aceptable.

2.1.6 Evaluación de los datos en la Escala Likert

Se elaboró un instrumento utilizando la escala de Likert, con preguntas cerradas, referidas a compromiso del docente con el estudiante, vicios que adquiere el estudiante durante sus años de estudio, tecnologías especializadas de la carrera, embarazo planificado o no planificado, asignación de la carrera a través del Senescyt, Bulling, entre otros; la intencionalidad de las preguntas planteadas fue valorar las opciones y determinar que factor es el más influyente en la deserción estudiantil aplicando la siguiente escala de valoración, como se puede observar en la tabla 5.

Tabla 5: Valores de la escala de likert.

Especificación	Valor	Simbología
No Influye	1	NI
Baja Influencia	2	BI
Media Influencia	3	MI
Alta Influencia	4	AI
Influye Totalmente	5	IT

Elaborado por: Investigador.

Una vez determinada la escala y cuyos resultados permiten un análisis descriptivo de los puntajes obtenidos en cada una de las variables para observar tendencias o variaciones en las preguntas aplicadas a los expertos.

Por último, las variables obtenidas al realizar las encuestas a los expertos fueron tabuladas, donde básicamente se realizó la clasificación de las variables con los

valores determinadas por la escala de Likert. Se procede a realizar el análisis de los de los valores totales que se presentan en los gráficos 12, 13, 14.

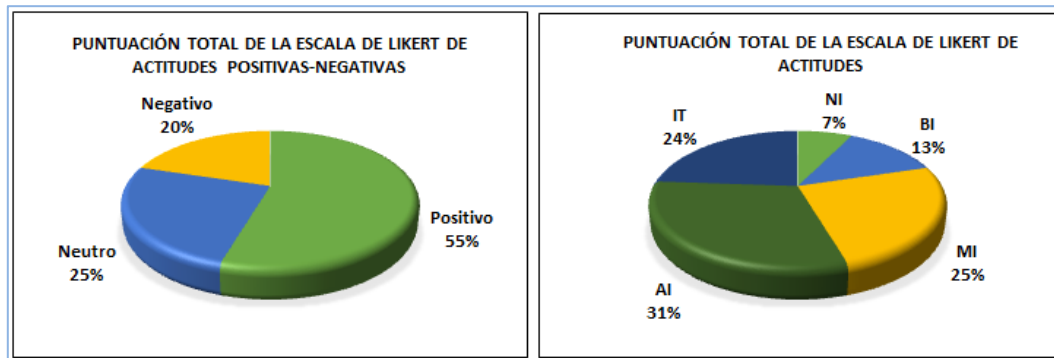


Gráfico 12: Distribución de frecuencia para los totales de datos recopilados.
Elaborado por: Investigador.

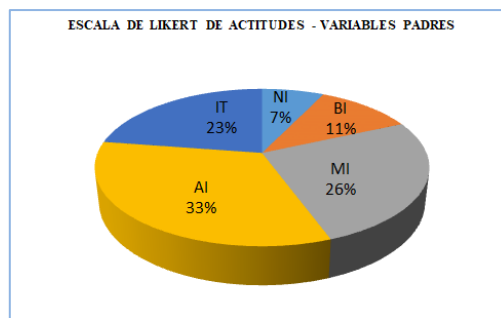


Gráfico 13: Distribución de frecuencia para las variables padres de deserción estudiantil.
Elaborado por: Investigador.

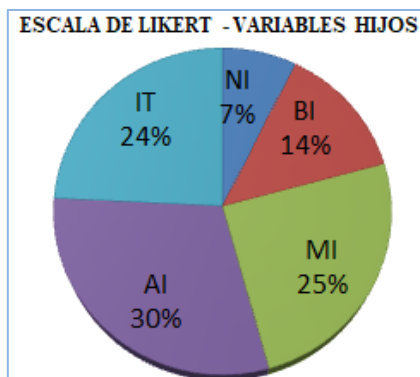


Gráfico 14: Distribución de frecuencia para las variables hijos, deserción estudiantil universitaria.
Elaborado por: Investigador.

2.1.7 Analística descriptiva de los datos (variables independientes)

Una vez realizado el análisis global de las variables padre e hijos, basadas en las puntuaciones obtenidas de escala de Likert aplicada a nuestro instrumento de investigación se puede observar que existen niveles de causalidad respecto a las respuestas de los expertos con la deserción en las universidades

Utilizando la herramienta Rstudio como lenguaje de programación “R” se procede a realizar el análisis descriptivo de los datos recolectados, con el instrumento de investigación aplicado a los expertos conocedores de la problemática. Este análisis permite determinar las variables padres utilizando la función *summary()* gráfico 15, la misma que nos retorna un resumen general sobre las variables del *data frame* (mínimo, máximo, media, mediana, primer y tercer cuartil) como se muestra en la tabla 6 y gráfico 15.

```
#####  
#Programa para análisis descriptivo de las variables padres**#  
#Edwin G. Chimba  
#####  
install.packages("xlsx")  
library(xlsx)  
#leer datos y asignar a la variable datosP  
datosP = read.xlsx("Procesor1.xlsx",1)  
View(datosP)  
#Resumen de los dataframe de variables padres|  
summary(datosP)  
#asignamos el resumen de los datos a un variable  
descriptivosPadres <- summary(datosP)  
#exportamos descriptivosPadres a un archivo .csv  
write.csv2(descriptivosPadres, file = "datosDesPadres.csv")
```

*Gráfico 15: Código del proceso descriptivo de las variables padres.
Elaborado por: Investigador.*

Tabla 6: Valores del análisis descriptivo de las variables padres.

VARIABLES PADRES							
V1	V2	V3	V4	V5	V6	V7	V8
Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:4.00	1 st Qu.:3.000	1 st Qu.:2.000	1 st Qu.:2.00	1 st Qu.:2.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000
Median :4.00	Median :4.000	Median :3.000	Median :3.00	Median :3.000	Median :4.000	Median :4.000	Median :4.000
Mean :4.19	Mean :3.233	Mean :3.043	Mean :3.04	Mean :2.933	Mean :3.491	Mean :3.426	Mean :3.696
3 rd Qu.:5.00	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.00	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000
Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
V9	V10	V11	V12	V13	V14	V15	V16
Min. :1.000	Min. :2.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:2.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:2.000	1 st Qu.:2.000	1 st Qu.:1.000	1 st Qu.:4.000	1 st Qu.:4.000
Median :3.000	Median :4.000	Median :4.000	Median :3.000	Median :4.000	Median :3.000	Median :5.000	Median :5.000
Mean :3.018	Mean :3.586	Mean :3.356	Mean :2.929	Mean :3.417	Mean :2.776	Mean :4.298	Mean :4.414
3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
V17	V18	V19	V20				
Min. :1.000	Min. :1.000	Min. :1.000	Min. :2.000				
1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:4.000				
Median :4.000	Median :4.000	Median :4.000	Median :4.000				
Mean :3.856	Mean :3.788	Mean :4.009	Mean :4.046				
3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:5.000				
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000				

Elaborado por: Investigador.

De la misma forma, en esta sección se realizará el análisis descriptivo de las variables hijos utilizando la función `summary()` que se presenta en el gráfico 16, la misma que nos retorna un resumen general sobre las variables del *data frame* (mínimo, máximo, media, mediana, primer y tercer cuartil) como se muestra en la tabla 7.

```
#####
#Programa para análisis descriptivo de las variables hijos##
#Edwin G. Chimba
#####
install.packages("xlsx")
library(xlsx)
#leer datos y asignar a la variable datosH
datosH = read.xlsx("Procesor1.xlsx",2)
view(datosH)
#Resumen de los dataframe de variables hijos
summary(datosH)
#asignamos el resumen de los datos a un variable
descriptivosHijos <- summary(datosH)
#exportamos descriptivosHijos a un archivo .csv
write.csv2(descriptivosHijos, file = "datosHijos.csv")
```

Gráfico 16: Código de análisis descriptivo de las variables hijos.

Elaborado por: Investigador.

Tabla 7: Valores del análisis descriptivo de las variables hijos.

VARIABLES HIJOS							
V1A	V1B	V1C	V2A	V2B	V2C	V2D	V2E
Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000
1 st Qu.:4.000	1 st Qu.:4.00	1 st Qu.:4.000	1 st Qu.:3.00	1 st Qu.:4.000	1 st Qu.:2.000	1 st Qu.:3.00	1 st Qu.:3.000
Median :4.000	Median :4.00	Median :5.000	Median :4.00	Median :4.000	Median :4.000	Median :4.00	Median :4.000
Mean :3.969	Mean :4.16	Mean :4.285	Mean :4.11	Mean :4.181	Mean :3.209	Mean :3.34	Mean :3.454
3 rd Qu.:5.000	3 rd Qu.:5.00	3 rd Qu.:5.000	3 rd Qu.:5.00	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:4.00	3 rd Qu.:4.000
Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000
V2F	V2G	V2H	V2I	V2J	V2K	V6A	V6B
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:1.000	1 st Qu.:1.00	1 st Qu.:2.000	1 st Qu.:2.000	1 st Qu.:1.000	1 st Qu.:4.000
Median :4.000	Median :3.000	Median :2.000	Median :2.00	Median :3.000	Median :2.000	Median :2.000	Median :5.000
Mean :3.405	Mean :3.331	Mean :1.963	Mean :2.16	Mean :2.929	Mean :2.402	Mean :2.353	Mean :4.175
3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:3.000	3 rd Qu.:3.00	3 rd Qu.:4.000	3 rd Qu.:3.000	3 rd Qu.:3.000	3 rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
V7A	V7B	V7C	V8A	V8B	V8C	V10A	V10B
Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:2.00	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000
Median :3.00	Median :4.000	Median :4.000	Median :3.000	Median :4.000	Median :4.000	Median :3.000	Median :4.000
Mean :2.81	Mean :3.408	Mean :3.975	Mean :3.353	Mean :3.675	Mean :3.834	Mean :3.242	Mean :3.515
3 rd Qu.:3.00	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:4.000
Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
V10C	V10D	V11A	V11B	V11C	V11D	V11E	V12A
Min. :1.000	Min. :2.000	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:2.000	1 st Qu.:3.00	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:2.000	1 st Qu.:2.000
Median :3.000	Median :3.000	Median :3.000	Median :3.00	Median :3.000	Median :3.000	Median :3.000	Median :4.000
Mean :3.451	Mean :3.347	Mean :2.887	Mean :3.11	Mean :3.313	Mean :3.288	Mean :2.865	Mean :3.242
3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:3.000	3 rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
V12B	V12C	V13A	V13B	V13C	V14A	V14B	V18A
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:2.000	1 st Qu.:1.000	1 st Qu.:4.000	1 st Qu.:2.000	1 st Qu.:2.00	1 st Qu.:1.000	1 st Qu.:1.000	1 st Qu.:3.000
Median :3.000	Median :2.000	Median :5.000	Median :4.000	Median :3.00	Median :3.000	Median :2.000	Median :3.000
Mean :2.966	Mean :2.681	Mean :4.393	Mean :3.393	Mean :2.96	Mean :2.911	Mean :2.282	Mean :3.466
3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:4.00	3 rd Qu.:4.000	3 rd Qu.:3.000	3 rd Qu.:4.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000
V18B	V18C	V18D	V18E	V18F	V18G	V18H	V18I
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000
Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.000
Mean :3.571	Mean :3.693	Mean :3.773	Mean :3.859	Mean :3.862	Mean :3.828	Mean :3.819	Mean :3.822
3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
V19A	V19B	V19C	V19D	V19E	V19F	V19G	V19H
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000
1 st Qu.:3.000	1 st Qu.:4.000	1 st Qu.:2.000	1 st Qu.:3.000	1 st Qu.:4.000	1 st Qu.:3.00	1 st Qu.:3.000	1 st Qu.:3.000
Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.00	Median :4.000	Median :4.000
Mean :3.788	Mean :4.172	Mean :3.273	Mean :3.807	Mean :4.021	Mean :4.11	Mean :4.034	Mean :3.592
3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:5.00	3 rd Qu.:5.000	3 rd Qu.:4.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000
V19I	V19J	V19K	V19L	V19M	V19N	V20A	V20B
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000	Min. :1.000
1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:2.250	1 st Qu.:3.000	1 st Qu.:3.00	1 st Qu.:3.000	1 st Qu.:4.000
Median :4.000	Median :4.000	Median :3.000	Median :3.000	Median :4.000	Median :4.00	Median :4.000	Median :4.000
Mean :3.525	Mean :3.994	Mean :3.617	Mean :3.417	Mean :3.807	Mean :3.66	Mean :3.853	Mean :4.018
3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:5.00	3 rd Qu.:4.000	3 rd Qu.:5.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000
V20C	V20D	V20E	V20F	V20G			
Min. :1.000	Min. :1.000	Min. :1.000	Min. :2.000	Min. :1.000			
1 st Qu.:3.000	1 st Qu.:4.000	1 st Qu.:3.000	1 st Qu.:3.000	1 st Qu.:3.000			
Median :4.000	Median :4.000	Median :4.000	Median :4.000	Median :4.000			
Mean :3.764	Mean :4.147	Mean :3.626	Mean :3.632	Mean :3.917			
3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:4.000	3 rd Qu.:5.000	3 rd Qu.:5.000			
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000			

Elaborado por: Investigador.

Utilizando al máximo las bondades de la herramienta Rstudio y con la ayuda de la estadística descriptiva procedemos con la aplicación de la desviación estándar y la media a las variables padres y variables hijos de nuestro dataset, donde se mide la dispersión que existe en nuestro conjunto de datos en relación con su media se puede visualizar en las tablas 8 y tabla 9.

Tabla 8: Valores de la media y desviación estándar de las variables padres.

Variable Padres (Media y Desviación Estándar)					
variable	mean	std_dev	variable	mean	std_dev
V1	4,19	0,856	V11	3,356	1,077
V2	3,233	1,167	V12	2,929	1,22
V3	3,043	0,982	V13	3,417	1,196
V4	3,04	1,399	V14	2,776	1,327
V5	2,933	1,085	V15	4,298	1,038
V6	3,491	0,894	V16	4,414	0,875
V7	3,426	0,995	V17	3,856	1,187
V8	3,696	0,933	V18	3,788	1,044
V9	3,018	1,294	V19	4,009	0,854
V10	3,586	0,654	V20	4,046	0,797

Elaborado por: Investigador.

Tabla 9: Valores de la media y desviación estándar de las variables hijos.

Variable Hijos (Media y Desviación Estándar)								
variable	mean	std_dev	variable	mean	std_dev	variable	mean	std_dev
V1A	3,969	1,019	V10B	3,515	0,803	V18H	3,819	1,067
V1B	4,16	0,869	V10C	3,451	0,703	V18I	3,822	1,114
V1C	4,285	0,892	V10D	3,347	0,822	V19A	3,788	0,926
V2A	4,11	0,884	V11A	2,887	0,994	V19B	4,172	0,831
V2B	4,181	0,881	V11B	3,11	1,008	V19C	3,273	1,108
V2C	3,209	1,171	V11C	3,313	1,198	V19D	3,807	0,972
V2D	3,34	1,247	V11D	3,288	1,264	V19E	4,021	0,832
V2E	3,454	1,266	V11E	2,865	1,131	V19F	4,11	0,951
V2F	3,405	1,253	V12A	3,242	1,501	V19G	4,034	1,03
V2G	3,331	1,354	V12B	2,966	1,234	V19H	3,592	1,062
V2H	1,963	1,202	V12C	2,681	1,487	V19I	3,525	1,013
V2I	2,16	1,15	V13A	4,393	0,891	V19J	3,994	1,038
V2J	2,929	1,139	V13B	3,393	1,202	V19K	3,617	1,089
V2K	2,402	1,185	V13C	2,96	1,251	V19L	3,417	1,108
V6A	2,353	1,223	V14A	2,911	1,391	V19M	3,807	1,127
V6B	4,175	1,16	V14B	2,282	1,275	V19N	3,66	1,288
V7A	2,81	1,142	V18A	3,466	0,943	V20A	3,853	0,857
V7B	3,408	1,012	V18B	3,571	0,877	V20B	4,018	0,796
V7C	3,975	1,176	V18C	3,693	0,947	V20C	3,764	0,885
V8A	3,353	1,002	V18D	3,773	1,083	V20D	4,147	0,843
V8B	3,675	0,97	V18E	3,859	1,086	V20E	3,626	0,983
V8C	3,834	1,234	V18F	3,862	1,149	V20F	3,632	1,035
V10A	3,242	0,859	V18G	3,828	1,099	V20G	3,917	0,946

Elaborado por: Investigador.

2.2 Análisis de correlación

Una vez realizado el análisis de confiabilidad de los datos, así como también un análisis descriptivo exhaustivo de cada una de las variables almacenados en los dos dataset, se procede a determinar las variables tanto padres como hijos más influyentes para el modelo de factores de deserción estudiantil en la Universidad Técnica de Cotopaxi.

2.2.1 Determinación de las variables independientes más influyentes

Integrando los paquetes MASS, ggplot2 y corrplot de Rstudio se realiza el análisis y selección de las variables padres más influyentes en la deserción estudiantil tal como se muestra en el gráfico 17.

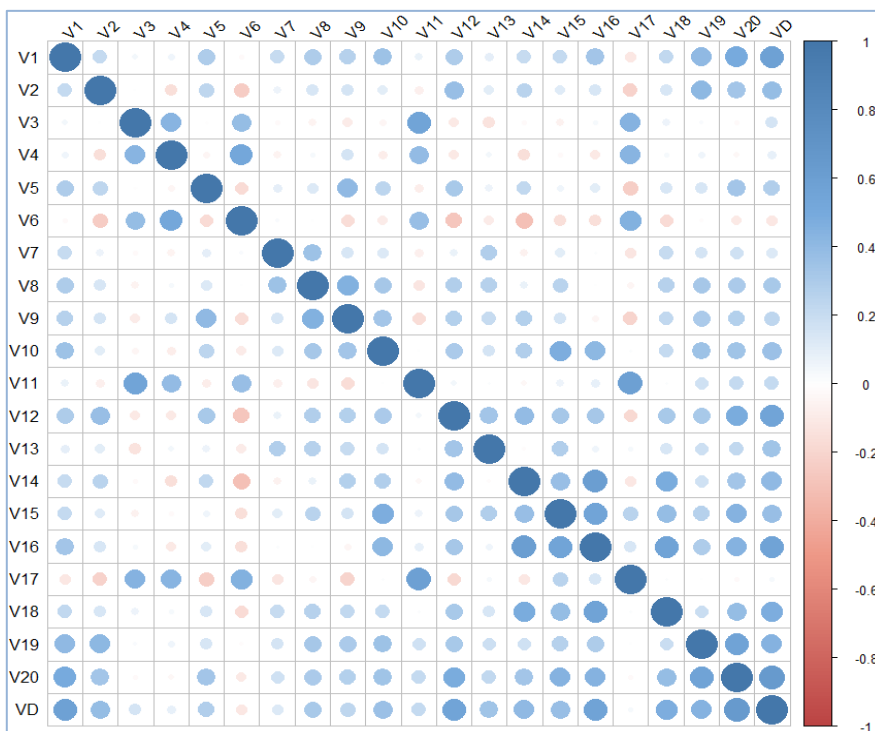


Gráfico 17: Análisis de correlación de las variables padres de deserción.
Elaborado por: Investigador.

El gráfico 18 presenta los factores padres más influyentes en la deserción estudiantil en el caso de estudio analizado.

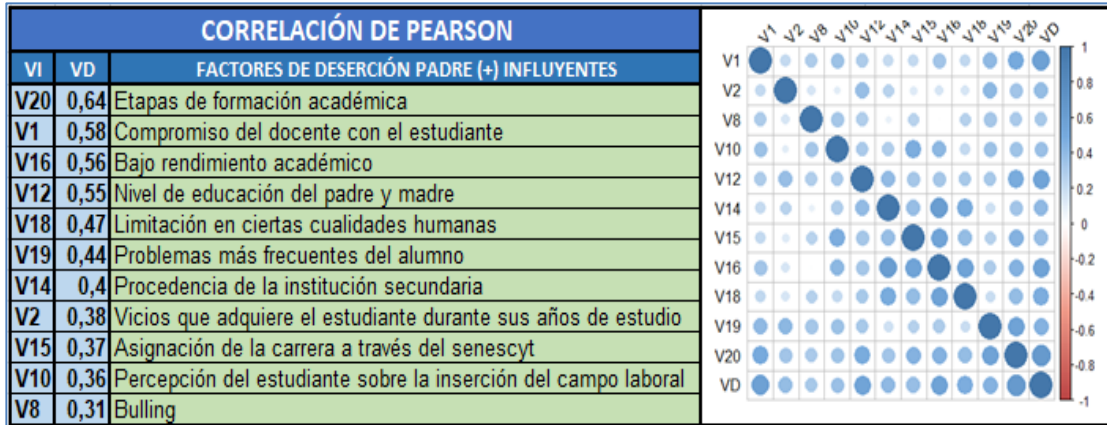


Gráfico 18: Factores de deserción padres más influyentes.
Elaborado por: Investigador.

A continuación, procedemos al análisis el dataset donde están almacenadas las variables hijos, haciendo uso de la herramienta Rstudio y aplicando la función corrplot() que nos ayuda a determinar la asociación y correlación entre variables de manera gráfica para poder interpretar de mejor manera los resultados, como se puede observar en el gráfico 19.

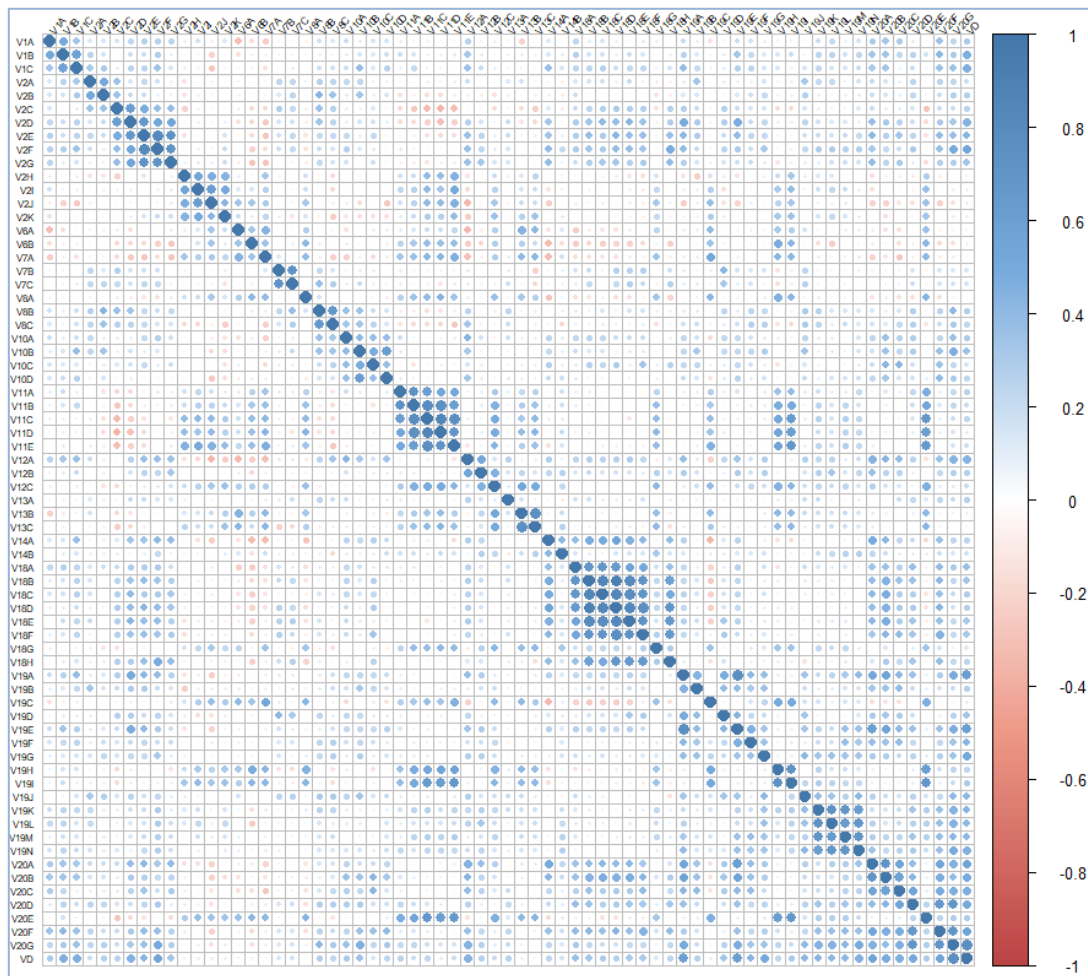


Gráfico 19: Análisis de correlación de las variables Hijos.
Elaborado por: Investigador.

Se puede determinar la relación existente entre las variables consideradas como hijos dentro del proceso experimental, la cual compone de una matriz de 15 indicadores, tal como se muestra en el gráfico 20.

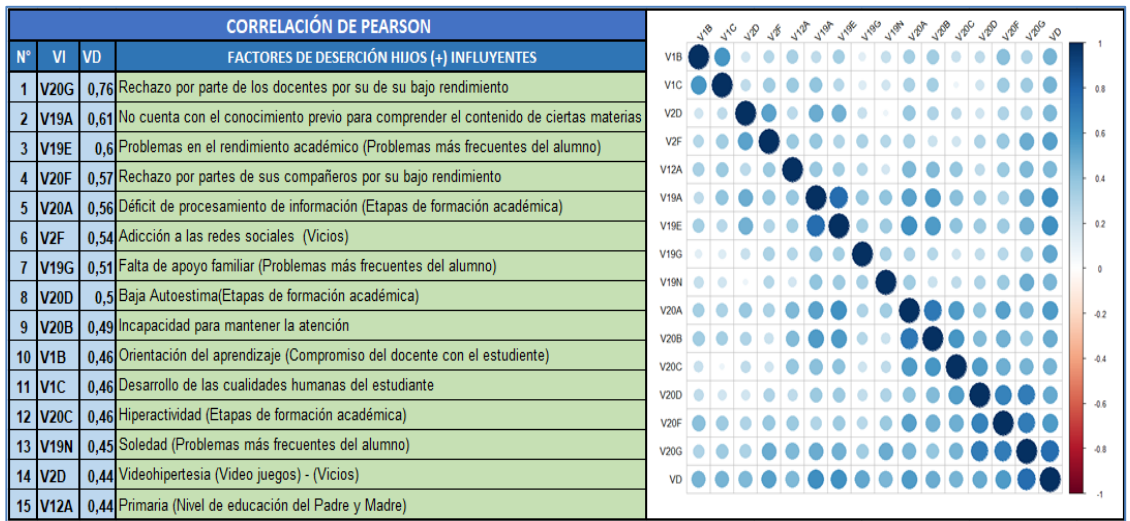


Gráfico 20: Factores de deserción hijos más influyentes.
 Elaborado por: Investigador.

Una vez que se ha establecido la confiabilidad de los datos y revisando la data de general, determinamos que existe relación entre los datos analizados, todos los datos se encuentran en las dos escalada definidas (0 y 1) aplicando la función `cor()` de R procedemos a determinar la correlación de la matriz mencionada, obteniendo un resultado como se muestra en el gráfico 21.

CORRELACIONES																						
Método	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	VD	
V1	pearson	1	0,21	0,04	0,05	0,29	-0,03	0,2	0,29	0,26	0,35	0,07	0,29	0,09	0,2	0,21	0,33	-0,12	0,22	0,4	0,49	0,58
V2	pearson	0,21	1	0	-0,16	0,23	-0,25	0,06	0,15	0,16	0,1	-0,08	0,37	0,1	0,25	0,11	0,14	-0,22	0,15	0,41	0,33	0,38
V3	pearson	0,04	0	1	0,43	-0,01	0,38	-0,03	-0,06	-0,1	-0,05	0,55	-0,11	-0,14	-0,03	-0,07	0,03	0,44	0,06	0,02	-0,03	0,16
V4	pearson	0,05	-0,16	0,43	1	-0,05	0,52	-0,06	0,03	0,16	-0,09	0,39	-0,11	0,04	-0,16	-0,03	-0,11	0,43	0,03	0,05	-0,04	0,08
V5	pearson	0,29	0,23	-0,01	-0,05	1	-0,18	0,09	0,13	0,4	0,24	-0,09	0,31	0,06	0,22	0,05	0,1	-0,24	0,15	0,15	0,33	0,28
V6	pearson	-0,03	-0,25	0,38	0,52	-0,18	1	0,02	0	-0,17	-0,1	0,37	-0,28	-0,1	-0,31	-0,16	-0,16	0,45	-0,18	-0,02	-0,11	-0,12
V7	pearson	0,2	0,06	-0,03	-0,06	0,09	0,02	1	0,35	0,15	0,13	-0,08	0,07	0,28	-0,07	0,1	0,01	-0,13	0,2	0,16	0,17	0,13
V8	pearson	0,29	0,15	-0,06	0,03	0,13	0	0,35	1	0,45	0,32	-0,13	0,28	0,26	0,07	0,25	0	-0,05	0,26	0,32	0,3	0,31
V9	pearson	0,26	0,16	-0,1	0,16	0,4	-0,17	0,15	0,45	1	0,33	-0,17	0,27	0,2	0,27	0,16	-0,05	-0,21	0,22	0,3	0,27	0,23
V10	pearson	0,35	0,1	-0,05	-0,09	0,24	-0,1	0,13	0,32	0,33	1	0	0,3	0,16	0,28	0,47	0,41	0,01	0,21	0,35	0,34	0,36
V11	pearson	0,07	-0,08	0,55	0,39	-0,09	0,37	-0,08	-0,13	-0,17	0	1	0,04	0,01	-0,04	0,06	0,08	0,59	0,01	0,17	0,21	0,21
V12	pearson	0,29	0,37	-0,11	-0,11	0,31	-0,28	0,07	0,28	0,27	0,3	0,04	1	0,33	0,39	0,32	0,32	-0,19	0,31	0,31	0,48	0,55
V13	pearson	0,09	0,1	-0,14	0,04	0,06	-0,1	0,28	0,26	0,2	0,16	0,01	0,33	1	-0,02	0,28	0,05	0,03	0,15	0,18	0,22	0,34
V14	pearson	0,2	0,25	-0,03	-0,16	0,22	-0,31	-0,07	0,07	0,27	0,28	-0,04	0,39	-0,02	1	0,37	0,6	-0,12	0,48	0,17	0,33	0,4
V15	pearson	0,21	0,11	-0,07	-0,03	0,05	-0,16	0,1	0,25	0,16	0,47	0,06	0,32	0,28	0,37	1	0,55	0,25	0,38	0,26	0,44	0,37
V16	pearson	0,33	0,14	0,03	-0,11	0,1	-0,16	0,01	0	-0,05	0,41	0,08	0,32	0,05	0,6	0,55	1	0,15	0,56	0,29	0,44	0,56
V17	pearson	-0,12	-0,22	0,44	0,43	-0,24	0,45	-0,13	-0,05	-0,21	0,01	0,59	-0,19	0,03	-0,12	0,25	0,15	1	0,01	0	-0,03	0,03
V18	pearson	0,22	0,15	0,06	0,03	0,15	-0,18	0,2	0,26	0,22	0,21	0,01	0,31	0,15	0,48	0,38	0,56	0,01	1	0,19	0,38	0,47
V19	pearson	0,4	0,41	0,02	0,05	0,15	-0,02	0,16	0,32	0,3	0,35	0,17	0,31	0,18	0,17	0,26	0,29	0	0,19	1	0,57	0,44
V20	pearson	0,49	0,33	-0,03	-0,04	0,33	-0,11	0,17	0,3	0,27	0,34	0,21	0,48	0,22	0,33	0,44	0,44	-0,03	0,38	0,57	1	0,64
VD	pearson	0,58	0,38	0,16	0,08	0,28	-0,12	0,13	0,31	0,23	0,36	0,21	0,55	0,34	0,4	0,37	0,56	0,03	0,47	0,44	0,64	1

Gráfico 21: Correlación de variables padres.
 Elaborado por: Investigador.

De la misma manera utilizando la herramienta SPSS, se procede a evaluar la correlación bivariada de la matriz compuesta por las variables hijos, teniendo como resultado el gráfico 22.

	V1A	V1B	V1C	V2A	V2B	V2C	V2D	V2
V1A	1	,514**	,328**	,128*	,139*	,137*	,219**	,26
Sig. (bilateral)		,000	,000	,021	,012	,013	,000	
N	326	326	326	326	326	326	326	
V1B	,514**	1	,575**	,250**	,165**	-,020	,185**	,18
Sig. (bilateral)	,000		,000	,000	,003	,722	,001	
N	326	326	326	326	326	326	326	
V1C	,328**	,575**	1	,315**	,295**	,112*	,248**	,24
Sig. (bilateral)	,000	,000		,000	,000	,044	,000	
N	326	326	326	326	326	326	326	
V2A	,128*	,250**	,315**	1	,508**	,354**	,145**	,21
Sig. (bilateral)	,021	,000	,000		,000	,000	,009	
N	326	326	326	326	326	326	326	
V2B	,139*	,165**	,295**	,508**	1	,355**	,196**	,19
Sig. (bilateral)	,012	,003	,000	,000		,000	,000	
N	326	326	326	326	326	326	326	
V2C	,137*	-,020	,112*	,354**	,355**	1	,599**	,47
Sig. (bilateral)	,013	,722	,044	,000	,000		,000	
N	326	326	326	326	326	326	326	

** La correlación es significativa al nivel 0,01 (bilateral).

* La correlación es significativa al nivel 0,05 (bilateral).

Gráfico 22: Correlación de variables Hijos.

Elaborado por: Investigador.

Las pruebas realizadas a los dos conjuntos de datos aplicando la correlación de Pearson, ayuda a medir la relación que existe entre las variables almacenadas en cada dataset, la correlación toma un rango de valores de +1 a -1, un valor de 0 nos indica que no hay asociación entre las variables, un valor +0 indica que su correlación es positiva, mientras que un valor -0 indica que su correlación es negativa.

Correlación igual a cero: Cuando la correlación es igual a cero significa que no es posible determinar algún sentido de la covariación de las variables.

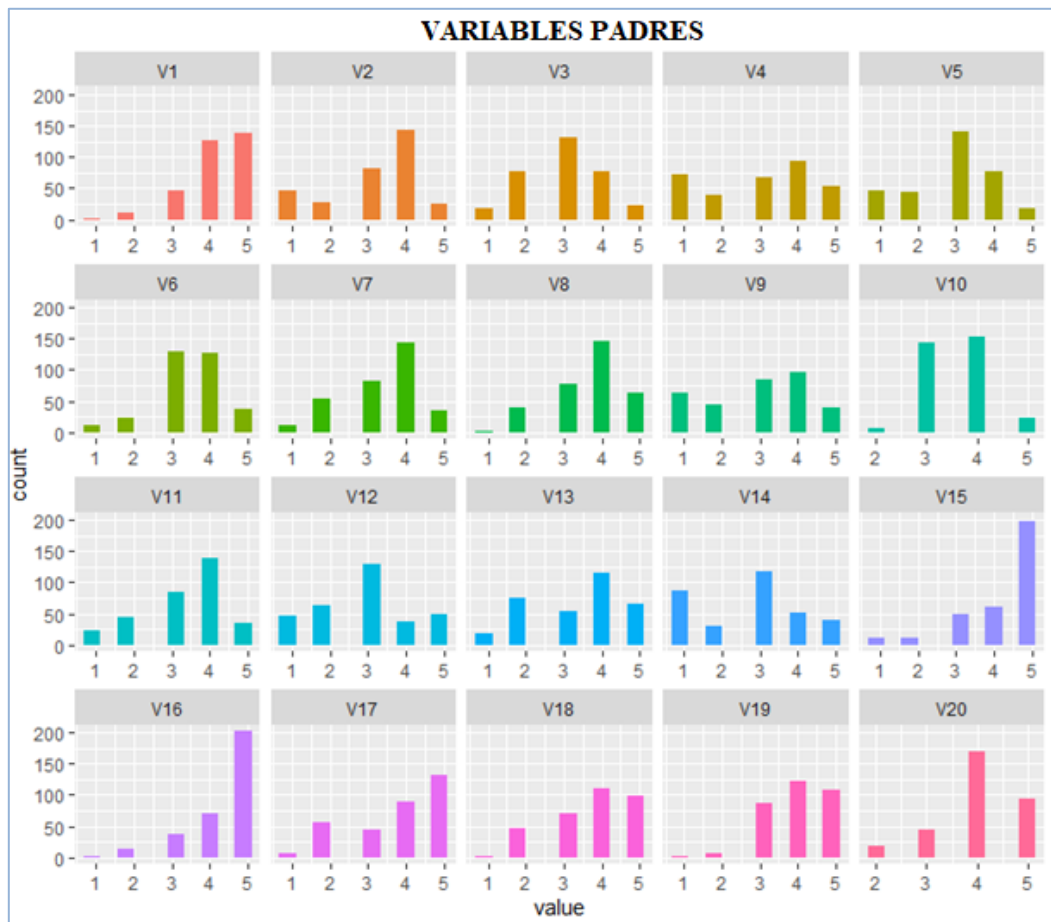
Correlación menor a cero: Significa que es negativa, es decir, que la relación de las variables es inversa.

Correlación mayor a cero: La correlación es positiva perfecta, es decir que las variables se correlacionan directamente.

Así, determinamos que en su mayoría las variables padres e hijos almacenada en cada dataset se correlacionan positivamente, convirtiéndose los datos seleccionados en un gran aporte para nuestra investigación.

2.3 Histograma de normalidad de datos

A continuación se procede a realizar la representación gráfica de histogramas de cada una de las variables como se muestra en los gráficos 23 y 24, para obtener estos resultados se utilizan las librerías `funModeling`, `tidyverse` y `Hmisc` conjuntamente con las funciones `profiling_num(datos)` con la que se pudo obtener la media y la desviación estándar, `plot_num(datos)` esta función representa gráficamente cada una de las variables numéricas de nuestra investigación, con el objetivo de visualizar que factor es el más influyente basado en la escala de Likert como se observa a continuación.



**Gráfico 23: Histogramas de las variables padres de deserción estudiantil.
Elaborado por: Investigador.**

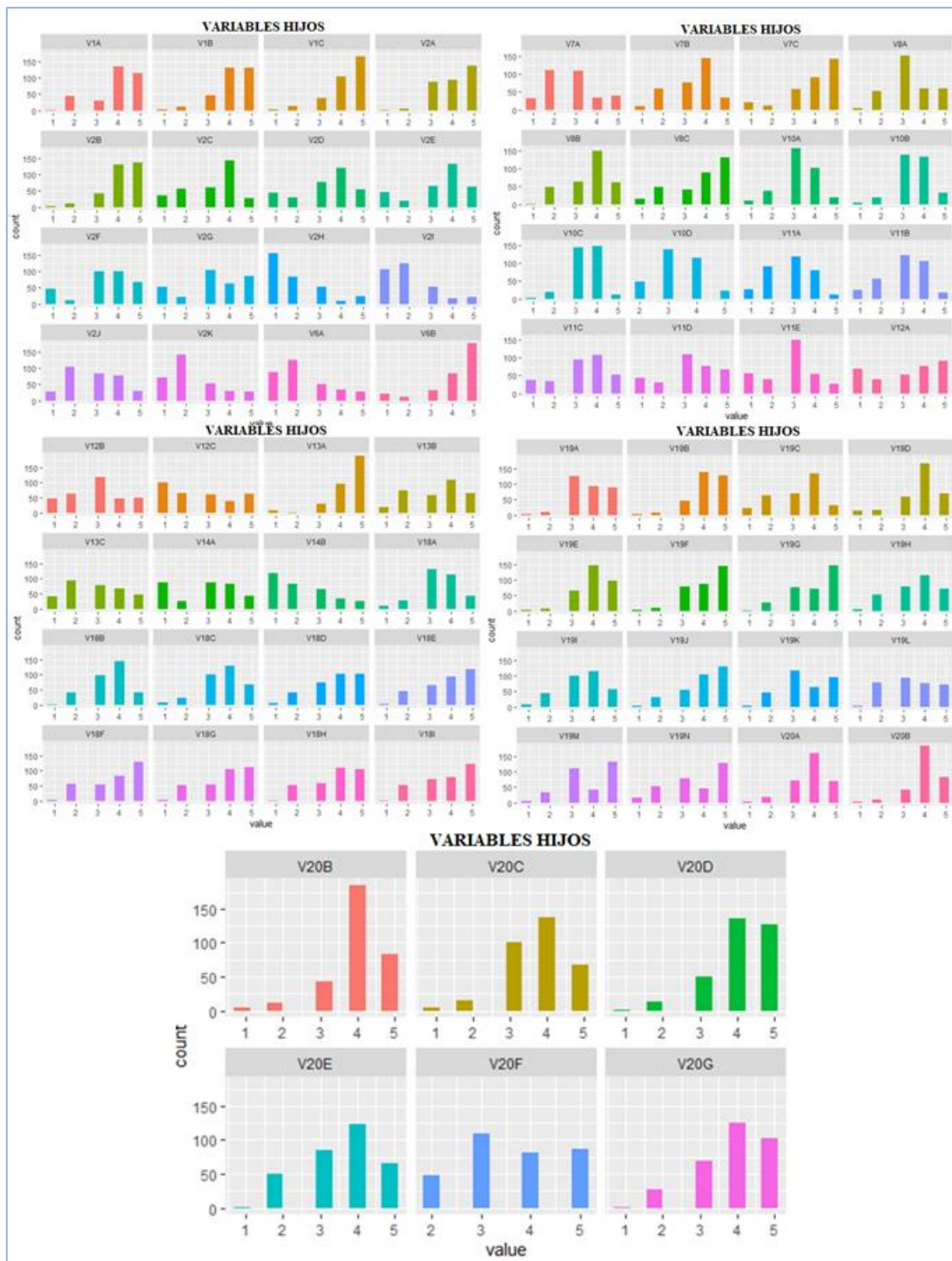


Gráfico 24: Histogramas de las variables hijos de deserción estudiantil.
 Elaborado por: Investigador.

El gran poder que tiene el lenguaje de programación R para el análisis de datos y aplicando correctamente las funciones antes mencionadas se puede las variables más representativas o influyentes que existen en nuestro conjunto de datos.

2.4 Modelo de deserción estudiantil universitaria

2.4.1 Modelo teórico

Definición de factores de deserción

Para determinar el modelo teórico haremos uso de todas las variables independientes y la variable dependiente deserción estudiantil universitaria seleccionada para este trabajo de investigación, determinadas a través de un proceso de validación de expertos y serán analizadas en un caso de estudio en la Universidad Técnica de Cotopaxi.

Para lo cual las variables independientes adoptaron el término de **variables padres** para hacer referencia a un factor de deserción estudiantil universitaria de manera general y el término **variables hijos** para hacer referencia a una componente del factor de deserción estudiantil, por ejemplo en el factor padre “**Vicios que adquiere el estudiante durante sus años de estudio**” tiene como factores hijos los vicios como: alcoholismo, drogadicción, ludopatía, adicción al ejercicio, adicción a la televisión, adicción al apego emocional, entre otros, como se describen a continuación en la tabla 10.

Tabla 10: Variables Independientes y dependiente para la deserción estudiantil en la U.T.C.

FACTORES DE DESERCIÓN PADRE	VARIABLE PADRE	VARIABLE HIJO	FACTORES DE DESERCIÓN HIJO
VARIABLES INDEPENDIENTES			
Compromiso del docente con el estudiante	V1	V1A	Formación docente
		V1B	Orientación del aprendizaje
		V1C	Desarrollo de las cualidades humanas del estudiante
Vicios que adquiere el estudiante durante sus años de estudio	V2	V2A	Alcoholismo
		V2B	Drogadicción
		V2C	Ludopatía
		V2D	Videohipertesia (Vídeo juegos)
		V2E	Ciberadicción (Internet)
		V2F	Adicción a las redes sociales
		V2G	Adicción a la Tecnología(Teléfonos, computadores, Tablet)
		V2H	Adicción al ejercicio
		V2I	Adicción a las compras
		V2J	Adicción al apego emocional
V2K	Adicción a la televisión		
Tecnologías especializadas de la carrera	V3	--	--
Compromiso económico del estudiante con la familia por ser hijo primogénito	V4	—	—
Disminución de categoría en el ranking institucional	V5	—	—
Embarazo planificado o no planificado	V6	V6A	Planificado
		V6B	No Planificado
Número de hijos del estudiante	V7	V7A	Entre 1 y 2 Hijos
		V7B	Entre 2 y 3 hijos
		V7C	Más de 3 hijos
Bulling	V8	V8A	De estudiante a estudiante
		V8B	De docente a estudiante
Machismo	V9	--	--
Percepción del estudiante sobre la inserción del campo laboral	V10	V10A	Competencias profesionales
		V10B	Situación económica actual
		V10C	Proyección económica de la profesión
		V10D	Creencias respecto a oportunidades laborales
Adaptación a las metodologías de formación universitaria	V11	V11A	Trabajo autónomo
		V11B	Trabajo cooperativo
		V11C	Trabajo grupal
		V11D	Trabajo colaborativo
		V11E	Utilización de tecnologías
Nivel de educación del padre y madre	V12	V12A	Primaria
		V12B	Secundaria
		V12C	Universitaria

Elaborado por: Investigador.

Tabla: 10 Variables Independientes y dependiente para la deserción estudiantil en la U.T.C. (Continuación)

VARIABLES INDEPENDIENTES			
Ingreso económico de los padres	V13	V13A	Menor a \$500
		V13B	Menor a \$1000
		V13C	Menor a \$1500
Asignación de la carrera a través de la Senescyt	V15	—	—
Bajo rendimiento académico	V16	--	--
Baja influencia de bienestar estudiantil	V17	--	--
Limitación en ciertas cualidades humanas	V18	V18A	Honestidad
		V18B	Adaptabilidad
		V18C	Persistencia
		V18D	Responsabilidad
		V18E	Disciplina
		V18F	Puntualidad
		V18G	Madurez
		V18H	Compromiso
		V18I	Actitud
Problemas más frecuentes del alumno	V19	V19A	La falta de conocimiento previo para comprender el contenido de ciertas materias
		V19B	Problemas económicos
		V19C	Problemas Sentimentales
		V19D	Dificultad en el aprendizaje
		V19E	Problemas en el rendimiento académico
		V19F	No se siente motivado con sus estudios
		V19G	Falta de apoyo familiar
		V19H	Grandes cargas de trabajo académico
		V19I	Problemas de adaptación con sus compañeros
		V19J	Problemas de Salud
		V19K	Ansiedad
		V19L	Timidez
		V19M	Depresión
		V19N	Soledad
Etapas de formación académica	V20	V20A	Déficit de procesamiento de información
		V20B	Incapacidad para mantener la atención
		V20C	Hiperactividad
		V20D	Baja Autoestima
		V20E	Problema para expresar los sentimientos
		V20F	Rechazo por partes de sus compañeros a causa del bajo rendimiento académico
		V20G	Rechazo por parte de los docentes por su bajo rendimiento
VARIABLE DEPENDIENTE			
DESERCIÓN ESTUDIANTIL	VD	--	--

Elaborado por: Investigador.

2.4.2 Regresión lineal múltiple

Para Montero [83] la regresión lineal múltiple trata de ajustar modelos lineales entre una variable dependiente y más de una variable independiente, esto quiere decir que el valor de una variable dependerá de la influencia o correlación de una tercera. Para generar el modelo se procedió a introducir todas las variables (89 variables) existentes en nuestro dataset en la herramienta Rstudio. Al ejecutar la función `summary()` a nuestro modelo, los valores de los coeficientes obtenidos, en este caso el *p-value* que es el elemento predictivo es representado como $\Pr(>|t|)$, la tabla 11 presenta los resultados obtenidos del procedimiento experimental.

Tabla 11: Regresión lineal.

	Estimate	Std Error	t value	Pr(> t)		Estimate	Std Error	t value	Pr(> t)
(Intercept)	-0,486	0,091	5,313	2,49E-07 ***	V12B	0,014	0,018	0,742	0,459077
V1	-0,119	0,125	0,95	0,342917	V12C	-0,087	0,034	2,584	0,010358 *
V1A	-0,215	0,039	5,513	9,21E-08 ***	V13	0,25	0,088	2,855	0,004681 **
V1B	0,521	0,111	4,677	4,90E-06 ***	V13A	-0,103	0,07	1,463	0,144923
V1C	-0,114	0,052	2,194	0,029219 *	V13B	-0,04	0,027	1,469	0,143237
V2	0,18	0,06	2,982	0,003169 **	V13C	-0,138	0,04	3,434	0,000702 ***
V2A	0,209	0,059	3,562	0,000446 ***	V14	0,353	0,088	3,986	8,97E-05 ***
V2B	-0,163	0,038	4,237	3,25E-05 ***	V14A	-0,374	0,083	4,492	1,10E-05 ***
V2C	0,175	0,029	6,054	5,53E-09 ***	V14B	0,051	0,017	2,943	0,003577 **
V2D	0,14	0,039	3,584	0,000411 ***	V15	-0,038	0,057	0,658	0,51093
V2E	0,037	0,056	0,672	0,50203	V16	0,278	0,07	3,965	9,74E-05 ***
V2F	-0,046	0,067	0,692	0,489345	V17	0,139	0,038	3,625	0,000354 ***
V2G	-0,113	0,03	3,745	0,000227 ***	V18	0,538	0,081	6,648	2,05E-10 ***
V2H	-0,095	0,031	3,079	0,002324 **	V18A	-0,075	0,035	2,12	0,035032 *
V2I	0,026	0,029	0,901	0,368524	V18B	-0,167	0,053	3,164	0,001763 **
V2J	0,002	0,026	0,092	0,926862	V18C	-0,253	0,048	5,253	3,34E-07 ***
V2K	0,135	0,025	5,432	1,38E-07 ***	V18D	-0,008	0,071	0,117	0,906639
V3	0,023	0,025	0,904	0,366703	V18E	0,195	0,056	3,473	0,000612 ***
V4	0,063	0,023	2,805	0,005452 **	V18F	0,008	0,038	0,204	0,83822
V5	0,014	0,025	0,553	0,580853	V18G	0,001	0,018	0,064	0,949307
V6	0,488	0,114	4,283	2,68E-05 ***	V18H	0,122	0,06	2,024	0,044122 *
V6A	-0,011	0,029	0,363	0,716691	V18I	-0,45	0,07	6,4	8,27E-10 ***
V6B	-0,5	0,113	4,44	1,38E-05 ***	V19	0,379	0,055	6,879	5,37E-11 ***
V7	0,163	0,088	1,86	0,064146 *	V19A	-0,11	0,067	1,637	0,103066
V7A	0,031	0,029	1,057	0,291629	V19B	-0,019	0,036	0,523	0,601405
V7B	0,093	0,035	2,619	0,009401 **	V19C	-0,045	0,019	2,41	0,016734 *
V7C	-0,291	0,076	3,811	0,000177 ***	V19D	-0,144	0,044	3,242	0,001359 **
V8	0,488	0,056	8,73	4,73E-16 ***	V19E	-0,499	0,061	8,202	1,52E-14 ***
V8A	-0,015	0,02	0,733	0,464466	V19F	0,043	0,033	1,3	0,194877
V8B	-0,253	0,051	4,946	1,44E-06 ***	V19G	-0,097	0,032	2,996	0,003031 **
V8C	0,144	0,036	3,999	8,51E-05 ***	V19H	-0,072	0,029	2,529	0,012084 *
V9	0,02	0,037	0,536	0,592355	V19I	0,038	0,027	1,404	0,161682
V10	-0,397	0,068	5,807	2,04E-08 ***	V19J	0,148	0,03	4,938	1,50E-06 ***

Elaborado por: Investigador.

Tabla 11: Regresión lineal. (Continuación)

	Estimate	Std Error	t value	Pr(> t)		Estimate	Std Error	t value	Pr(> t)		
V10B	0,182	0,056	3,233	0,001402	**	V19L	0,13	0,031	4,129	5,05E-05	***
V10C	-0,103	0,067	1,545	0,123691		V19M	-0,158	0,035	4,501	1,06E-05	***
V10D	0,07	0,04	1,774	0,077333	,	V19N	-0,058	0,026	2,193	0,029273	*
V11	-0,328	0,073	4,471	1,21E-05	***	V20	0,045	0,068	0,656	0,512176	
V11A	0,004	0,024	0,173	0,862934		V20A	0,265	0,043	6,211	2,35E-09	***
V11B	0,02	0,035	0,579	0,562823		V20B	0,301	0,057	5,233	3,68E-07	***
V11C	0,34	0,064	5,284	2,87E-07	***	V20C	-0,013	0,051	0,251	0,802194	
V11D	0,019	0,032	0,602	0,54754		V20D	-0,225	0,037	6,022	6,55E-09	***
V11E	0,143	0,033	4,294	2,57E-05	***	V20E	-0,073	0,026	2,768	0,006085	**
V12	-0,153	0,073	2,099	0,03689	*	V20F	0,068	0,044	1,524	0,128834	
V12A	0,247	0,062	4,017	7,92E-05	***	V20G	0,291	0,055	5,325	2,34E-07	***

Elaborado por: Investigador.

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.07763 on 236 degrees of freedom
Multiple R-squared:  0.9619,    Adjusted R-squared:  0.9475
F-statistic: 66.94 on 89 and 236 DF,  p-value: < 2.2e-16
    
```

Gráfico 25: Resultados del modelo de regresión lineal.

Elaborado por: Investigador.

Como se puede observar en el gráfico 25, una vez generado y corrido el modelo inicial con todas las variables introducidas como predictores, el resultado muestra un R-cuadrado alto correspondiente a 0.9619, esto quiere decir que el modelo es capaz de explicar el 96% de la viabilidad observada en la deserción estudiantil universitaria, el *p-value* obtenido es muy pequeño y corresponde a 2,2e-16, lo que indica la relación de causalidad entre las variables independientes y la variable dependiente deserción estudiantil universitaria (VD).

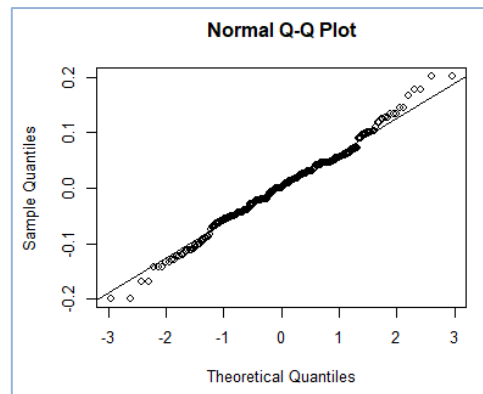


Gráfico 26: Representación gráfica del modelo original de regresión lineal.

Elaborado por: Investigador.

Modelo ajustado con variables significativas de deserción estudiantil universitaria

Para ajustar nuestro modelo procedemos a ejecutar la función **step()** al conjunto de datos del modelo inicial, obteniendo así 31 variables que aportan significativamente al modelo propuesto de deserción estudiantil universitaria, como se muestra en la tabla 12.

Tabla 12: Modelo de regresión lineal ajustado.

	Estimate	Error	t value	Pr(> t)		Estimate	Error	t value	Pr(> t)		
(Intercept)	-0,3293	0,05814	-5,664	3,59E-08	***	V13A	-0,10638	0,02406	-4,421	1,39E-05	***
V1A	-0,09492	0,03208	-2,959	0,003344	**	V16	0,16453	0,04341	3,791	0,000183	***
V1B	0,27619	0,0427	6,468	4,27E-10	***	V17	0,08356	0,02812	2,972	0,00321	**
V2A	0,2621	0,06005	4,365	1,78E-05	***	V18	0,39676	0,06977	5,687	3,19E-08	***
V2C	0,14946	0,02609	5,729	2,55E-08	***	V18C	-0,25319	0,04893	-5,175	4,30E-07	***
V2D	0,14235	0,03092	4,604	6,24E-06	***	V18E	0,14985	0,05025	2,982	0,003109	**
V2G	-0,06606	0,02321	-2,847	0,004736	**	V18I	-0,26464	0,04736	-5,588	5,33E-08	***
V2K	0,07845	0,02399	3,27	0,001207	**	V19	0,36642	0,04535	8,08	1,81E-14	***
V8	0,23717	0,0561	4,227	3,18E-05	***	V19E	-0,46266	0,04804	-9,631	< 2e-16	***
V8B	-0,15621	0,04173	-3,744	0,000219	***	V19L	0,07869	0,02547	3,09	0,002201	**
V8C	0,09409	0,03572	2,634	0,008894	**	V19M	-0,0958	0,02597	-3,689	0,000269	***
V10	-0,18125	0,03802	-4,767	2,98E-06	***	V20A	0,21802	0,03585	6,081	3,80E-09	***
V10A	0,1171	0,02935	3,99	8,40E-05	***	V20B	0,17637	0,04101	4,301	2,33E-05	***
V11	-0,30178	0,05497	-5,49	8,87E-08	***	V20D	-0,18264	0,03374	-5,413	1,31E-07	***
V11C	0,28789	0,05026	5,728	2,56E-08	***	V20G	0,31659	0,04321	7,327	2,40E-12	***
V11E	0,15184	0,0296	5,13	5,35E-07	***	V12A	0,12954	0,02484	5,214	3,54E-07	***

Elaborado por: Investigador.

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1233 on 287 degrees of freedom
Multiple R-squared:  0.8831,    Adjusted R-squared:  0.8677
F-statistic: 57.08 on 38 and 287 DF,  p-value: < 2.2e-16
```

Gráfico 27: Visualización de los resultados del modelo ajustado de regresión lineal.

Elaborado por: Investigador.

Como se puede observar en el gráfico 27, una vez generado el modelo ajustado seleccionando las variables más significativas, con base a determinar aquellas significativas estadísticamente, utilizando el 95% de confianza de la técnica. Como resultado se obtiene 31 variables predictores con valores R-cuadrado 0.8677, esto quiere decir que el modelo es capaz de explicar el 87% de la viabilidad observada en

la deserción estudiantil universitaria, el *p-value* obtenido corresponde a $2,2e-16$ lo que permite establecer el nivel de causalidad de las variables analizadas con la deserción.

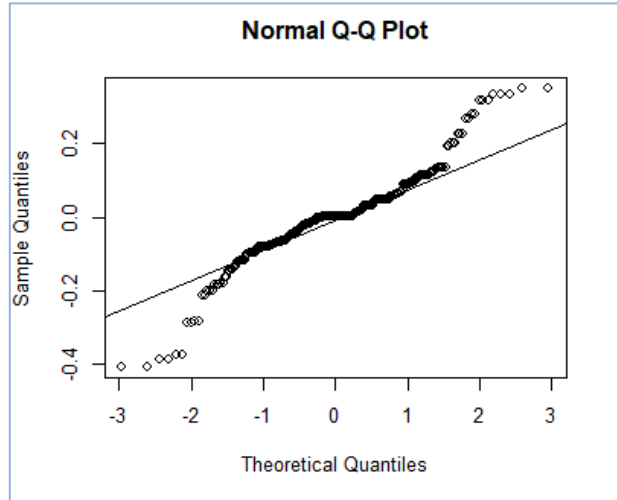


Gráfico 28: Resultados del modelo ajustado de regresión lineal.
Elaborado por: Investigador.

El modelo de regresión lineal también supone de los residuos, para lo cual la gráfica de la Normal Q-Q que permite compara la distribución de los residuos con la distribución norma teórica, esto quiere decir, que si los residuos tiene una distribución normal no debe apartarse de la diagonal, como se puede observar en el gráfico 28.

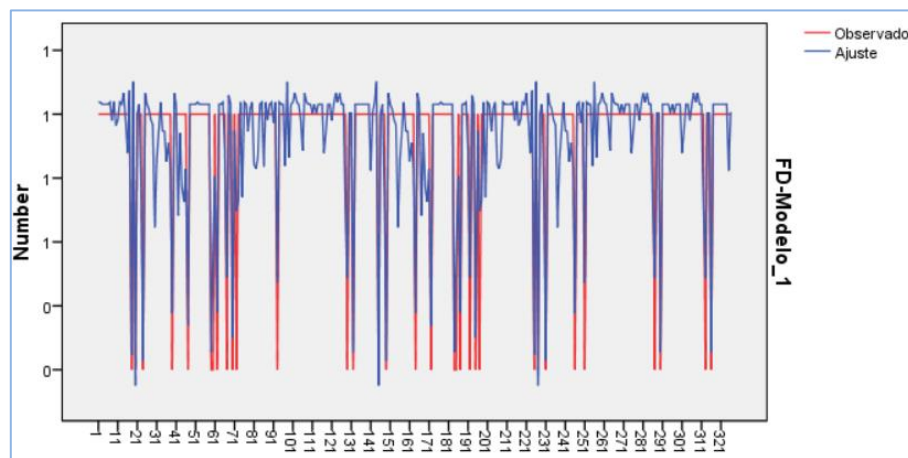


Gráfico 29: Ajuste del modelo.
Elaborado por: Investigador.

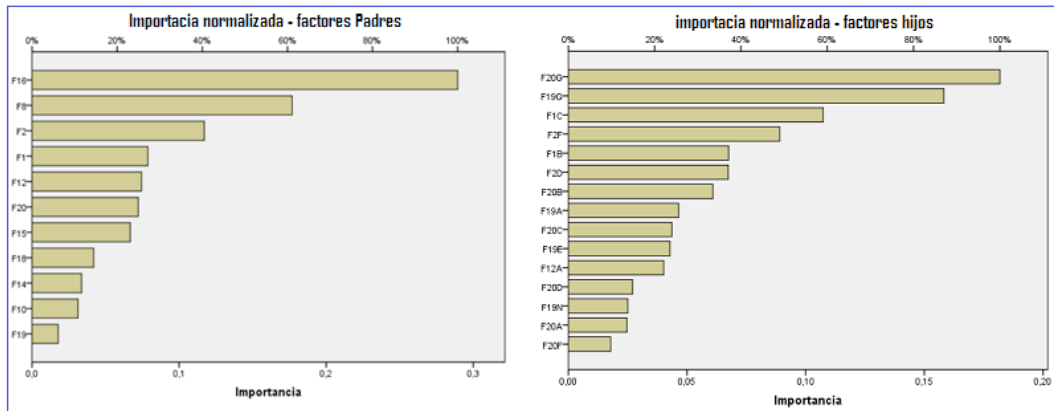


Gráfico 30: Factores más significativos del modelo.
 Elaborado por: Investigador.

A continuación, se presenta el cumplimiento de los supuestos del modelo de regresión lineal como son: linealidad, independencia, homocedasticidad, normalidad y no colinealidad, ejecutando estos comandos en Rstudio `par(mfrow = c(2,2))` `plot(modelo_Mejorado)`, demostrando así que el modelo es válido, como se muestra en la imagen gráfico 31.

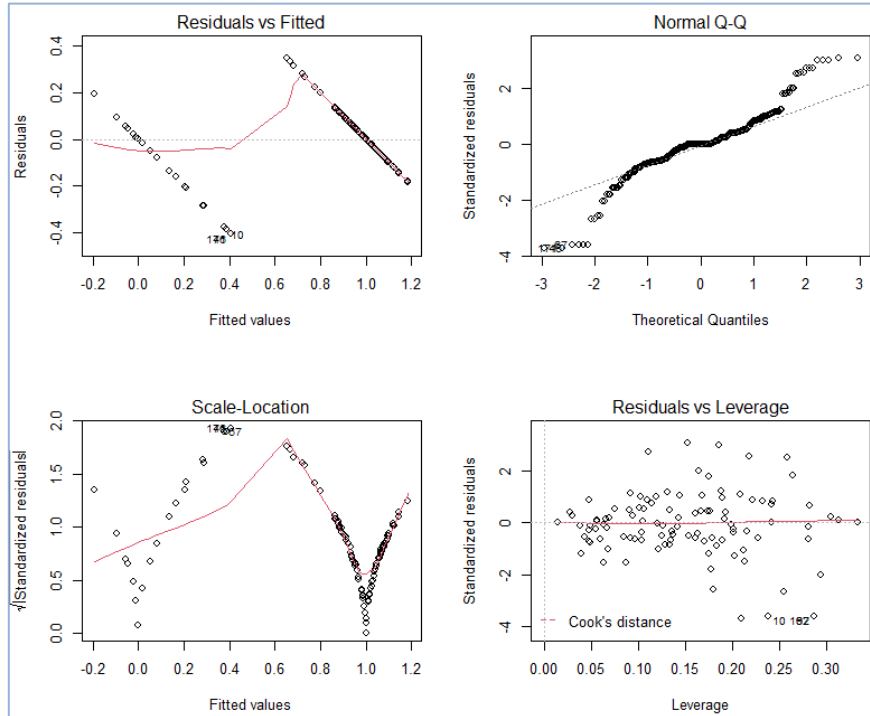


Gráfico 31: Comprobación de supuestos del modelo de regresión lineal
 Elaborado por: Investigador.

2.4.3 Modelo conceptual

El modelo conceptual se construye con los factores padres - hijos de deserción estudiantil, tomados como base a los 31 factores, identificados como los más influyentes que causan la deserción estudiantil en la Universidad Técnica de Cotopaxi, como se observa en el gráfico 32.

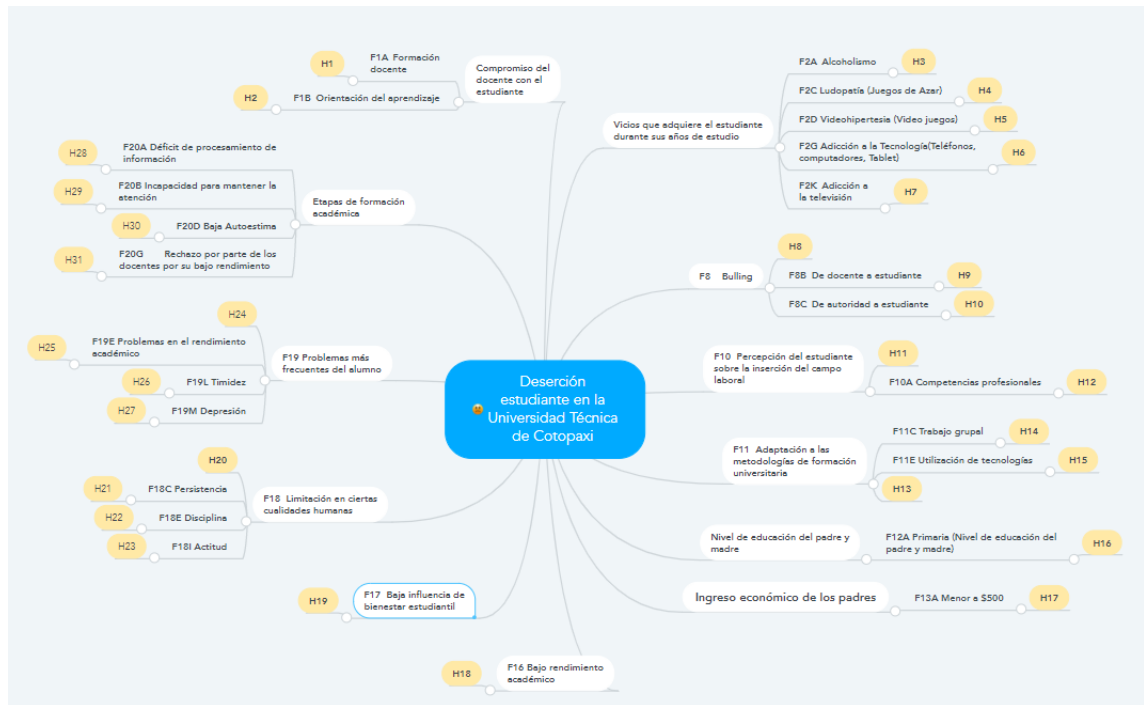


Gráfico 32: Modelo conceptual de deserción estudiantil para la U.T.C.
Elaborado por: Investigador

2.4.4 Modelo lógico

Identificando el conjunto de los factores de deserción estudiantil más influyentes dentro de nuestra investigación, se obtuvo 31 factores que se presentan en la tabla 13.

Tabla 13: Factores más influyentes del modelo ajustado de deserción estudiantil.

Factor	Descripción	Factor	Descripción
F1A	Formación del docente	F13A	Ingreso económico de los padres menor a 500 USD
F1B	Orientación del aprendizaje	F16	Bajo rendimiento académico
F2A	Consumo de alcohol	F17	Limitada influencia de servicios de bienestar estudiantil
F 2C	Ludopatía (Juegos de Azar)	F18	Limitación cualidades humanas
F 2D	Video hipertesia (Videojuegos)	F18C	Desmotivación en la persistencia
F 2G	Adicción a la Tecnología(Teléfonos, computadores, Tablet)	F18E	Limitada disciplina
F2K	Adicción a la televisión	F18I	Limitada actitud de éxito
F8	Bulling	F19	Problemas personales
F8B	De estudiante a estudiante	F19E	Bajo rendimiento académico
F8C	De autoridad a estudiante	F19L	Timidez
F10	Percepción del estudiante sobre la inserción del campo laboral	F19M	Depresión
F10A	Competencias profesionales	F20A	Déficit de procesamiento de información
F11	Adaptación a las metodologías de formación universitaria	F20B	Incapacidad para mantener la atención
F11C	Trabajo grupal	F20D	Baja Autoestima
F11E	Conocimiento en el uso de tecnologías	F20G	Limitada atención por parte de los docentes por su bajo rendimiento
F12A	Nivel de educación del padre y madre Primaria		

Elaborado por: Investigador.

FDE-UTC-1: Formación docente.- Se refiere a la formación del docente como formación instrumental que se realiza a través donde adquieren conocimiento y habilidades didácticas para poner en práctica con los estudiantes [84], la falta de esta formación por parte del docente influyendo fuertemente en la deserción del estudiante.

Hipótesis H1: La formación académica docente influye en la deserción estudiantil universitaria.

FDE-UTC-2: Orientación del aprendizaje.- Se refiere a la orientación del aprendizaje por parte del docente, cuando el estudiante encuentra las clases más interesantes, genera un interés en el proceso de aprendizaje [85] creando así un

ambiente agradable para su formación académica, una limitación en la orientación por parte del docente influye fuertemente en la deserción del estudiante.

***Hipótesis H2:** La Orientación del aprendizaje por parte del docente influye en la deserción estudiantil universitaria.*

FDE-UTC-3: Alcoholismo.- Se refiere a que el alcoholismo como causa de deserción universitaria, como menciona Albarracín y Muñoz [86] el consumo de bebida alcohólicas en la jornada de clases representa un desprendimiento de sus responsabilidades como estudiantes, influyendo fuertemente en la deserción universitaria.

***Hipótesis H3:** El consumo de alcohol influye en la deserción estudiantil universitaria.*

FDE-UTC-4: Ludopatía.- Se refiere a la adicción para los juegos de azar creando así una dependencia en el estudiante y dejando un lado sus metas académicas, provocando la deserción universitaria [87].

***Hipótesis H4:** La Ludopatía (inclinación a Juegos de Azar) influye en la deserción estudiantil en la Universidad Técnica de Cotopaxi.*

FDE-UTC-5: Videohipertesia.- Se refiere a la adicción de los videos juegos, como manifiesta Churra [88] que el uso de videojuegos en el horario de clase, provoca la desconcentración en el proceso de aprendizaje, así también el uso de videos juegos en el tiempo libre provoca un descuido en su formación académica perdiendo el interés en los estudios.

***Hipótesis H5:** La Videohipertesia (Videojuegos) influye en la deserción estudiantil en la Universidad Técnica de Cotopaxi.*

FDE-UTC-6: Adicción a la Tecnología.- Se refiere a adicción que crea en el estudiante el uso aparatos tecnológicos como teléfonos, computadores, tabletas entre otros, provocando así la falta de interés en los estudios.

***Hipótesis H6:** La Adicción a la Tecnología (Teléfonos, computadores, Tablet) influye en la deserción estudiantil en la Universidad Técnica de Cotopaxi.*

FDE-UTC-7: Adicción a la televisión.- Se refiere a adicción de la televisión por parte del estudiante en sus horas libres, creando un completo desinterés en la responsabilidad con sus estudios. Maya, Aguilar, Zamora y Barron [89] manifiesta que la adicción de la televisión es uno de los factores que afecta a la atención que debe tener el alumno para cumplir con sus actividades académicas.

***Hipótesis H7:** La Adicción a la televisión influye en la deserción estudiantil.*

FDE-UTC-8: Bulling.- Se refiere al Bulling que se crea dentro y fuera de las instituciones, Albán [90] comenta que si el estudiante es agredido físicamente o emocional genera problemas en el proceso de aprendizaje, influyendo fuertemente en la deserción universitaria del estudiante.

***Hipótesis H8:** El Bulling influye en la deserción estudiantil.*

FDE-UTC-9: Bulling de docente a estudiante.- Se refiere a la agresión física o emocional del docente hacia el estudiante, provocando así el desánimo de continuar en sus estudios e influye fuertemente en la deserción universitaria.

***Hipótesis H9:** El Bulling de docente a estudiante influye en la deserción estudiantil.*

FDE-UTC-10: Bulling de autoridad a estudiante.- Se refiere a la agresión física o emocional de una autoridad de la institución educativa hacia el estudiante, influyendo fuertemente en la deserción universitaria del estudiante.

Hipótesis H10: *El Bulling de autoridad a estudiante influye en la deserción estudiantil.*

FDE-UTC-11: Percepción del estudiante sobre la inserción del campo laboral.- Se refiere a la percepción del estudiante sobre la inserción del campo laboral una vez que el estudiante llegue a concluir sus estudios, Albán [90] sostiene que el estudiante crea una idea o se proyecta en un futuro de cómo lo aprendido en las aulas de dicha institución pueden aportar en campo laboral, influyendo fuertemente en la deserción universitaria del estudiante con esta proyección.

Hipótesis H11: *La percepción del estudiante sobre la inserción del campo laboral influye en la deserción estudiantil.*

FDE-UTC-12: Competencias profesionales.- Se refiere a las competencias profesionales que alcanza el estudiante al finalizar sus estudios, esto demanda de la incertidumbre que tiene el estudiante a creer que no está capacitado adecuadamente para asumir un cargo en una institución pública o privada.

Hipótesis H12: *Las Competencias profesionales influyen en la deserción estudiantil.*

FDE-UTC-13: Adaptación a las metodologías de formación universitaria.- Se refiere a la adaptación a las metodologías de formación universitaria por parte del estudiante, Arancibia y Rodríguez [91] manifiesta que los jóvenes deben adaptarse a un sistema de educación más complejos, pues la transición del nivel secundaria a la universidad en algunos casos existirá desnivelación es sus conocimientos.

Hipótesis H13: *La adaptación a las metodologías de formación universitaria influye en la deserción estudiantil.*

FDE-UTC-14: Limitada adaptación al trabajo grupal.- Se refiere a trabajo grupal que debe realizar el estudiante como complemento de sus actividades académicas, es implica que el estudiante debe convivir con varios estudiantes de diferentes costumbres, niveles socioeconómicos, etnias, entre otros, la falta de empatía con estos grupos provocaría limitación en la formación académica cuando el estudiante no puede adaptarse a grupos de trabajo sintiéndose excluido del grupo de sus compañeros.

Hipótesis H14: *La limitada adaptación al trabajo grupal influye en la deserción estudiantil universitaria.*

FDE-UTC-15: Utilización de tecnologías.- Se refiere a la utilización de tecnologías para el proceso de aprendizaje por parte del estudiante, el manejo o utilización de tecnologías es se hace muy necesaria para complementar lo aprendido en clases, por lo que la no utilización de las tecnologías existentes influye en la deserción universitaria.

Hipótesis H15: *La nula utilización de tecnologías influye en la deserción estudiantil.*

FDE-UTC-16: Nivel de educación del padre y madre - Primaria.- Se refiere al nivel de educación del padre y madre en este caso hablamos que los padres solo culminaron la primaria, influyendo altamente en la deserción estudiantil universitaria [92].

Hipótesis H16: *El nivel de educación del padre y madre influye en la deserción estudiantil universitaria.*

FDE-UTC-17: Ingreso económico de los padres.- Se refiere al ingreso económico de los padres, el poder adquisitivo en nuestra sociedad influye en la determinación de recursos que otorgan los padres a sus hijos para su permanencia en la universidad [93].

Hipótesis H17: *El Ingreso económico de los padres influye en la deserción estudiantil universitaria.*

FDE-UTC-18: Bajo rendimiento académico.- Se refiere a que el estudiante no consigue seguir el nivel de aprendizaje de reto de estudiantes de la clase y esta situación es una constante en ciclo tras ciclo [94].

Hipótesis H18: *El bajo rendimiento académico influye en la deserción estudiantil.*

FDE-UTC-19: Limitados servicios de bienestar estudiantil.- Se refiere a que Bienestar Estudiantil debe velar por el adecuado desarrollo del estudiante dentro de la universidad, interactuando con el estudiante contantemente para motivarle a que finalice su carrera universitaria. De no ser así, influye altamente den la deserción estudiantil universitaria.

Hipótesis H19: *La Baja influencia de bienestar estudiantil influye en la deserción estudiantil.*

FDE-UTC-20: Limitación en ciertas cualidades humanas.- Se refiere a que el estudiante no tiene la misma capacidad de aprendizaje que el resto de estudiantes, dando lugar así a una diferencia notoria en el rendimiento académico, basado en diferentes factores internos como pueden ser adaptabilidad al grupo, honestidad consigo mismo, disciplina, puntualidad, Madurez entre otros [95].

Hipótesis H20: *La Limitación en ciertas cualidades humanas influye en la deserción estudiantil.*

FDE-UTC-21: Persistencia.- Se refiere a una de las cualidades humanas que nos ayuda intentar una y otra vez pese a los fracasos que hayamos tenido, con esto quiere decir que la persistencia ayuda a la superación de los obstáculos [96].

Hipótesis H21: *La Persistencia influye en la deserción estudiantil.*

FDE-UTC-22: Disciplina.- Hace referencia a otra cualidad humana, la misma que nos ayuda a mantenernos firmes en nuestro objetivo a pesar de las circunstancias adversas que se presenten, la carencia de esta cualidad influirá altamente en la deserción universitaria [96].

Hipótesis H22: *La Disciplina influye en la deserción estudiantil.*

FDE-UTC-23: Actitud.- Se refiere a otra cualidad humana que hacer referencia a la voluntad que posee el estudiante para culminar sus estudios [96].

Hipótesis H23: *La Actitud influye en la deserción estudiantil.*

FDE-UTC-24: Problemas más frecuentes del alumno.- Se refiere a que el estudiante comienza a tener varios problemas dentro de la institución como los problemas económicos, problemas sentimentales, dificultad de aprendizaje, motivación por sus estudios, problemas de adaptabilidad con sus compañeros, falta de apoyo familiar, problemas de salud, entre otros [97].

Hipótesis H24: *Los Problemas más frecuentes del alumno influyen en la deserción estudiantil.*

FDE-UTC-25: Problemas en el rendimiento académico.- Se refiere a que el estudiante comienza a tener varios problemas dentro de la institución como los problemas económicos, problemas sentimentales, dificultad de aprendizaje,

motivación por sus estudios, problemas de adaptabilidad con sus compañeros, falta de apoyo familiar, problemas de salud, entre otros [94].

Hipótesis H25: *Los Problemas en el rendimiento académico influyen en la deserción estudiantil.*

FDE-UTC-26: Timidez.- Se refiere a la timidez del estudiante en el aula de clases. Castaño y Gallego [98] plantea que la timidez social es uno de los factores importantes a la hora de investigar la deserción estudiantil, puesto que el saber que piensa, siente y como actúa el individuo es importante para entablar un diálogo fluido, cosa que una persona tímida no va a poder realizar. Creando un gran problema al estudiante que posee esta cualidad.

Hipótesis H26: *La Timidez como cualidad personal influye en la deserción estudiantil.*

FDE-UTC-27: Depresión.- Se refiere a que la depresión de uno de los problemas más frecuentes del alumno que cursa el nivel superior, Agudelo [99] citando a Lozano 2003 manifiesta que la depresión tiene antecedentes como: personales, familiares, dificultades académicas, muerte de un ser querido, separación de los padres, diagnóstico de enfermedades graves, inestabilidad económica, son algunas razones por las que se ve afectado un estudiante en su estado de ánimo, por lo que su bajo rendimiento académico se vuelve más evidente.

Hipótesis H27: *La Depresión influye en la deserción estudiantil.*

FDE-UTC-28: Déficit de procesamiento de información.- Se refiere al déficit de procesamiento de información por parte del estudiante cuando está cursando el nivel superior, Pérez, Valenzuela, Díaz y González [100] manifiesta que los estudiantes se relaciona directamente con el aprendizaje cognitivo orientando su atención hacia la

comprensión de lo que está estudiando de no ser así disminuye su responsabilidad y muestra menor motivación para aprender.

Hipótesis H28: *El Déficit de procesamiento de información influye en la deserción estudiantil.*

FDE-UTC-29: Incapacidad para mantener la atención.- Se refiere a la incapacidad para mantener la atención en las horas clases por parte del estudiante, provocando así un desfase en el aprendizaje con el resto de estudiantes. Influyendo fuertemente en la deserción del estudiante [95].

Hipótesis H29: *La Incapacidad para mantener la atención influye en la deserción estudiantil.*

FDE-UTC-30: Baja Autoestima.- Se refiere a la baja autoestima que tenga el estudiante a la hora de asistir a la Universidad, al no sentirse seguro de poder concluir una carrera en una institución de tercer nivel. Influyendo fuertemente en la deserción del estudiante [98].

Hipótesis H30: *La Baja Autoestima influye en la deserción estudiantil.*

FDE-UTC-31: Rechazo por parte de los docentes por su bajo rendimiento.- Se refiere al rechazo por parte del docente por su bajo rendimiento académico, el docente puede ser muy enfático y no ser tolerante a falta del compromiso del estudiante en sus horas clases. Influyendo fuertemente en la deserción del estudiante de la Universidad.

Hipótesis H31: *El Rechazo por parte de los docentes por su bajo rendimiento influye en la deserción estudiantil.*

2.5 Conclusiones del Capítulo II

- Para la determinación de factores de deserción estudiantil universitaria, se procede a la recolección de información basada en las opiniones de expertos, haciendo uso de la encuesta como instrumento de investigación para el presente proyecto, las variables identificadas a través de este proceso sirvieron como variables de ingreso al modelo teórico propuesto.
- La tasa de respuesta de la opinión de expertos dejó como resultado un conjunto de datos con 326 observaciones y 89 variables que para el nuestro trabajo de investigación son factores de deserción estudiantil, se procede a realizar la validación del instrumento de investigación utilizando el Alfa de Cronbach teniendo como resultado un valor del 0.925 siendo mayor que el 0.8 considerado como mínimo aceptable, así dando validez a los datos recolectados.
- Del conjunto total de datos donde están almacenadas todas las variables de deserción estudiantil se procede a dividir en dos conjunto de datos compuestos por las variables padres y las variables hijos, aplicando el modelo de regresión lineal múltiple se terminan que de 89 variables de deserción 30 son las variables más significativas para el modelo teórico de deserción estudiantil.

CAPÍTULO III

MODELOS DE PREDICCIÓN DE FACTORES DE DESERCIÓN MEDIANTE MINERÍA DE DATOS

3.1 Proceso experimental para determinación de factores de deserción.

La minería de datos es una herramienta estratégica y clave para el desarrollo de este proyecto, donde se extraerá el conocimiento de los datos recolectados de los expertos. Para el proceso experimental se utiliza la técnica Knowledge Discovery in Databases (KDD) y la metodología (CRISP-CM) consideradas como técnicas adecuadas para el desarrollo de la investigación, estableciendo así un contexto mucho más rico e influyente en la elaboración de los modelos predictivos de deserción estudiantil, el proceso de la aplicación de las metodologías se presenta en la Gráfica 33.

De esta manera aplicando la metodología CRISP-CM que nos ayuda a estandarizar las fases aplicadas en la técnica para la minería de datos del presente proyecto teniendo como objetivo la determinación de factores que influyen en la deserción estudiantil universitaria, haciendo uso de redes neuronales perceptrón multicapa y Deep Learning pondremos en ejecución dicho proceso.

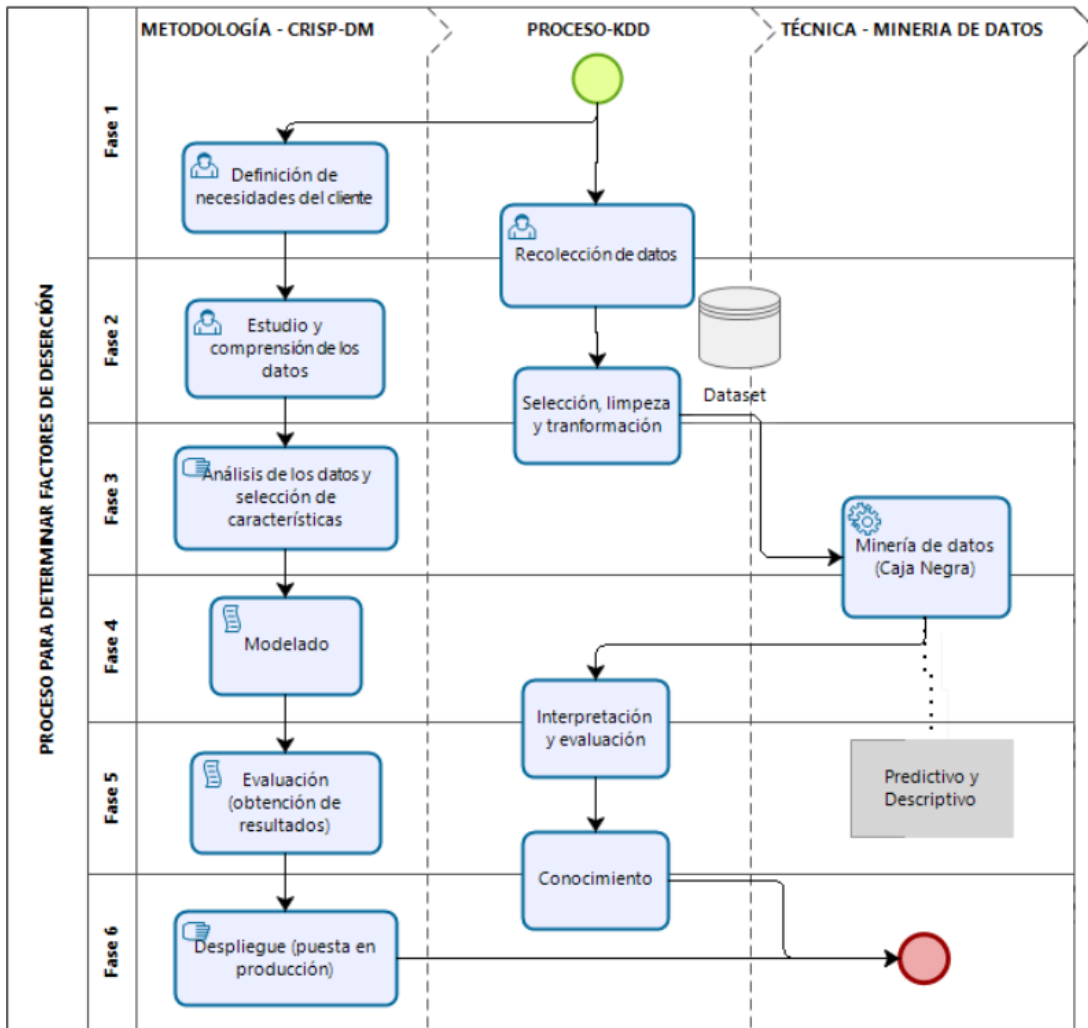
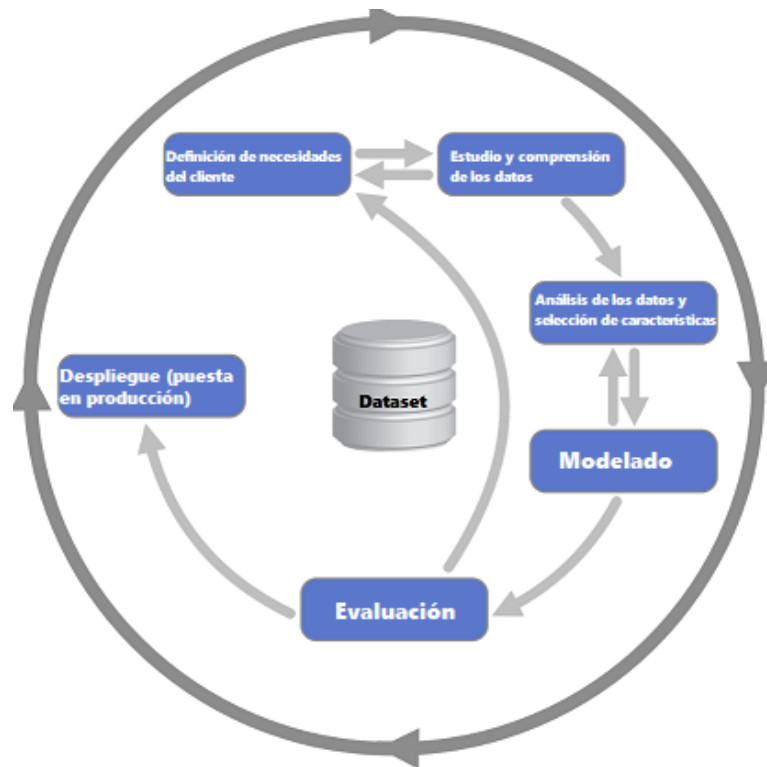


Gráfico 33: Proceso para la determinación de los modelos de predicción.
 Elaborado por: Investigador

El ciclo de vida de un proyecto de minería de datos consiste en 6 fases claramente identificadas en el gráfico 34. La secuencia de las fases no es rígida, se permite movimientos hacia adelante y hacia atrás entre diferentes etapas del proceso, el resultado de una fase determinará en particular que hay que hacer después. Las flechas indican las dependencias más importantes y frecuentes de esta metodología.



*Gráfico 34: Proceso para determinar factores de deserción.
Elaborado por: Investigador.*

3.1.1 Metodología

El proceso de minería de datos KDD - Knowledge Discovery in Databases [61] hace referencia al hallazgo del conocimiento en los datos, con la ayuda de la metodología CRISP-DM - Cross Industry Standard Process for Data Mining que básicamente normaliza el proceso de descubrimiento de conocimiento y al mismo tiempo implementa el ciclo de vida de un proyecto de análisis de datos, la metodología CRISP-DM diseñada para proyectos de minería de datos, tiene una gran ventaja, que las fases del proceso interactúan entre ellas de forma iterativa hasta que este finalice. El presente trabajo de investigación consiste en implementar un proceso predictor de factores de deserción estudiantil aplicando minería de datos basado en redes neuronales con el objeto de verificar que el algoritmo que tengan mejor comportamiento en la solución al problema planteado.

La investigación tiene como principal fuente de datos la información recopilada de las encuestas realizada a expertos de la Universidad Técnica de Cotopaxi, la misma que se realizó en noviembre del 2019, teniendo como resultado un dataset que será procesado para verificar cuáles de los factores (variables) más relevantes para el modelo de deserción estudiantil propuesto.

3.2 Fases para la implementación del proyecto

Para la elaboración de un proyecto de minería de datos, es de vital importancia cumplir a cabalidad cada una de las fases que demanda la metodología CRISP-DM, que básicamente nos ayuda a estandarizar dicho proyecto, siguiendo paso a paso cada uno de los conceptos establecidos a continuación:

3.2.1 Definición de necesidades del cliente (comprensión del negocio)

El análisis a ejecutarse es la determinación de factores de deserción estudiantil en la Universidad Técnica de Cotopaxi utilizando minería de datos, en esta fase se analizará los siguientes parámetros:

- **Determinar el objetivo del presente trabajo de investigación**, el objetivo es descubrir patrones de comportamiento en los alumnos para establecer indicadores de deserción estudiantil universitaria.
- **Evaluar la situación actual**, al tener la capacidad de establecer factores que afectan negativamente a la deserción en las universidades se puede obtener una herramienta de apoyo a la toma de decisiones de las autoridades universitarias.
- **Establecimiento de los objetivos de la minería de datos**, uno de los objetivos principales de esta investigación es apoyarnos en la minería de

datos, aplicando técnicas, procesos, metodología y tecnologías que permiten explorar grandes bases de datos, donde se encontraran patrones repetitivos o factores de deserción estudiantil.

- **Generación de plan del proyecto**, una vez definida la metodología, se propone analizar el conjunto de datos recolectado de los expertos utilizando las herramientas Rstudio, SPSS y algoritmos que vienen embebidos en estas.

3.2.2 Estudio y comprensión de los datos

En la Universidad Técnica de Cotopaxi existen una gran cantidad de expertos, que tienen el conocimiento necesario para determinar los factores de deserción estudiantil dentro del plantel mencionado, para esta fase de la metodología aplicada que comprende del entendimiento de los datos es necesario comenzar por:


3.2.2.1 Identificación de los datos

Una vez identificado de manera clara y precisa el problema de nuestra investigación procedemos a seleccionar la población y muestra para el proyecto de investigación, que en nuestro caso son los expertos que encuentra inmersos directa o indirectamente con el problema de deserción estudiantil en la Universidad Técnica de Cotopaxi. Se seleccionaron los datos recolectados de las opiniones de 326 expertos, en donde constan de 89 variables o factores que influyen en la deserción estudiantil.


3.2.2.2 Recolección de datos

Una vez seleccionada la muestra con la que se va a trabajar dentro del proyecto, de acuerdo con nuestro problema de estudio e hipótesis, la siguiente etapa consiste en recolectar los datos pertinentes sobre los conceptos o variable. Recolectar los datos iniciales para la investigación implica elabora un plan detallado que nos conduzca a

reunir datos con un propósito específico, con la ayuda del instrumento de investigación como se puede observar en el gráfico 35.



UNIVERSIDAD
TÉCNICA DE
COTOPAXI



POSGRADO

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 personas que tengan experiencia en temas relacionadas 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 llenan el cuestionario. La sección 2 corresponde a preguntas que determinan la influencia de factores que inciden en la deserción, y la sección 3 contiene preguntas que complementan el estudio.

GRACIAS POR SU COLABORACIÓN!!!

Gráfico 35: Encuesta para expertos.
Elaborado por: Investigador.

Aplicando correctamente el instrumento de investigación, se obtiene datos muy valiosos, los mismos que ha sido almacenado en Dataset, como se muestra en el gráfico 36, cumpliendo con una de las etapas de la fase inicial de la minería de datos.

2019/11/14 11:29:03 a.Á m. GMT-5,""Si","Si","Entre 1 y 2","La Matriz","Tutor","Pública","Analista de sistemas","Femenino","40","Ecuador","La Tacunga","5","5","5","5","4","3","3","3","2","2","2","2","1","2","1","5","2","3","4","3","4","4","2","4","4","4","4","3","3","3","3","La edad","5","5","5","4","4","1","1","5","3","4","4","4","5","5","4","4","4","													
A	B	C	D	E	F	G	H	I	J	K	L	M	
2019/11/14 11:20:10 a.Á m. GMT-5,""Si","Si","Menor a 1","La Matriz","Docente en el Área de Educación","Pública","Ingeniero","Femenino","35","Ecuador","Latacunga","3","3"													
2019/11/14 11:23:28 a.Á m. GMT-5,""Si","Si","Entre 2 y 3","La Matriz","Docente en el Área de Educación","Privada","Ing de Sistemas","Masculino","36","Ecuador","Ibarra","5","5"													
2019/11/14 11:23:59 a.Á m. GMT-5,""Si","Si","Menor a 1","La Matriz","Docente en el Área de Educación","Pública","Ingeniera","Femenino","28","Ecuador","Latacunga","5","5"													
2019/11/14 11:29:03 a.Á m. GMT-5,""Si","Si","Entre 1 y 2","La Matriz","Tutor","Pública","Analista de sistemas","Femenino","40","Ecuador","La Tacunga","5","5","5","5","4","3"													

Gráfico 36: Dataset inicial - opiniones expertos.
Elaborado por: Investigador.

3.2.2.3 Limpieza y transformación de los datos

El siguiente proceso consiste en la selección, limpieza y transformación de los datos, donde podremos identificar si existe o no datos atípicos, evaluando la calidad los mismos, para descartar la inconsistencia de información recolectada tenemos que establecer su nivel de confiabilidad. El proceso de limpieza de datos empieza con la separación de las secciones de nuestro instrumento de investigación, claramente identificados en el capítulo anterior. Lo cual se presenta en el gráfico 37.

Si	Si	Entre 1 y 2	La Matriz	Tutor	Pública	Ing. En Sistemas	Masculino	45	Ecuador	Latacunga
Si	Si	Entre 1 y 2	La Matriz	Tutor	Pública	Ing. En Sistemas	Femenino	39	Ecuador	Latacunga
Si	Si	Entre 1 y 2	La Matriz	Docente en el Área de Edu	Pública	Ing. Mecánica	Femenino	37	Ecuador	Cotopaxi
Si	Si	Menor a 1	La Matriz	Tutor	Pública	Ing. Contabilidad	Femenino	26	Ecuador	Latacunga
Si	Si	Entre 1 y 2	La Matriz	Miembro del Equipo de Ge	Pública	Ing. Diseño Gráfi	Femenino	29	Ecuador	Latacunga
Si	Si	Entre 1 y 2	La Matriz	Tutor	Pública	Ing. En Sistemas	Masculino	45	Ecuador	Latacunga
Si	Si	Entre 2 y 3	La Matriz	Docente en el Área de Edu	Pública	Ing. En Sistemas	Femenino	28	Ecuador	Latacunga
Si	Si	Más de 4	La Matriz	Docente en el Área de Edu	Pública	Ing. Mecánica	Masculino	45	Ecuador	Latacunga
Si	Si	Entre 1 y 2	La Matriz	Decano o Subdecano de Fe	Pública	Ing. En Sistemas	Masculino	28	Ecuador	Latacunga
Si	Si	Menor a 1	La Matriz	Psicólogo(a)	Pública	Psicólogo	Masculino	56	Ecuador	Latacunga
Si	Si	Menor a 1	La Matriz	Docente en el Área de Edu	Pública	Ing. Mecánica	Masculino	40	Ecuador	Latacunga
Si	Si	Menor a 1	La Matriz	Tutor	Pública	Magister en Doc	Femenino	67	Ecuador	Latacunga
Si	Si	Entre 3 y 4	La Matriz	Docente en el Área de Edu	Pública	Ing Eléctrico	Masculino	33	Ecuador	Latacunga
Si	Si	Menor a 1	La Matriz	Director o Coordinador de	Pública	Docente Investig	Masculino	59	Ecuador	Latacunga
Si	Si	Menor a 1	La Matriz	Tutor	Pública	Ing. Diseño Gráfi	Masculino	30	Ecuador	Latacunga
Si	Si	Entre 2 y 3	La Matriz	Decano o Subdecano de Fe	Pública	Ing. En Sistemas	Femenino	42	Ecuador	Cotopaxi
Si	Si	Menor a 1	La Matriz	Docente en el Área de Edu	Pública	Ing. Mecánica	Femenino	35	Ecuador	Latacunga

Gráfico 37: Limpieza y transformación de los datos.
Elaborado por: Investigador.

De las secciones 2 y 3 que se trata de los factores de deserción, al ejecutar el proceso de extracción y limpieza de los datos se obtiene un dataset con la información más relevante desde el punto de vista de los expertos, este proceso permite obtener información altamente confiable convirtiéndose así en la materia prima para el desarrollo del presente proyecto, tal como se muestra en el gráfico 38.

Formación docente															
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Formación de	Orientación	Desarrollo de	Alcoholismo	Drogadicción	Ludopatía	Videohipterpes	Ciberadicción	Adicción a la	Adicción a la	Adicción al	Adicción a la	Adicción a la	Adicción a la	Adicción a la	Adicción a la
5	4	4	4	3	2	4	4	4	4	4	4	3	2	4	4
5	4	4	4	3	2	4	4	4	4	4	4	3	2	4	4
4	3	4	4	4	3	4	5	4	3	2	2	2	2	2	3
4	4	3	4	4	4	4	3	3	3	2	2	2	2	3	4
4	5	5	5	5	4	4	4	3	3	1	2	4	2	3	1

Gráfico 38: Datos recolectados
 Elaborado por: Investigador.

Una vez realizado el proceso de extracción y limpieza de los datos se determina que el instrumento de investigación arrojó 89 variables padres de deserción y una variable dependiente que es la deserción estudiantil universitaria.

Determinados ya los factores de deserción se procede a separar en dos bloques de datos, por un lado los factores padres (20 variables padres de deserción) y por otro lado los factores hijos (69 variables hijos de deserción). El objetivo de esta separación categórica es para una mejor comprensión de los datos dentro de nuestra investigación, como se puede observar en el gráfico 39.

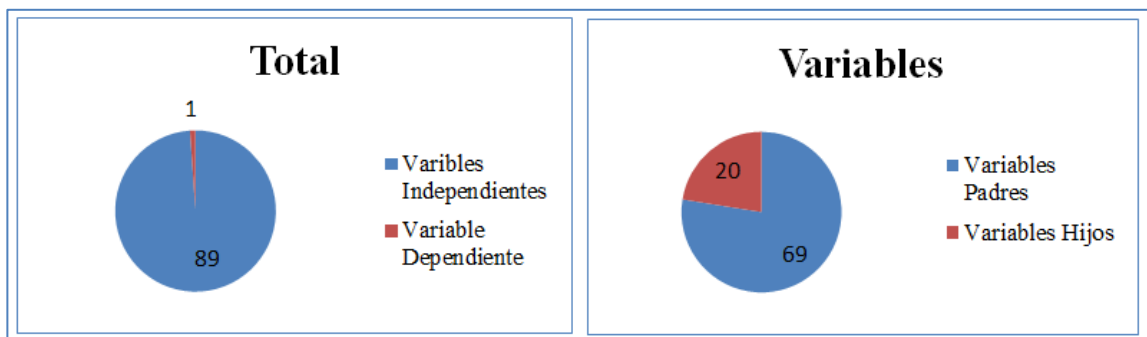


Gráfico 39: Total de factores de deserción.
 Elaborado por: Investigador.

Los datos recolectados están valorados por la escala de Likert del 1 al 5, por lo que en esta fase de transformación de datos se procede a codificar a datos binarios, el resultado del procesamiento de datos se presenta en los gráficos 40 y 41, concluyendo

exitosamente la fase de transformación de los datos obteniendo así un conjunto de datos depurada y lista para ser sometida las técnicas de minería de datos.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V12A	V12B	V12C	V13	V14	V15	V16	V17	V18	V19	V20	VD
319	1	1	0	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0
320	1	1	0	0	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1	1	1	1
321	1	1	0	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0
322	1	1	0	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0
323	1	1	0	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0
324	1	1	0	0	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0
325	1	1	0	0	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1	0	1	1
326	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1

Gráfico 40: Matriz de variables padres procesados.
Elaborado por: Investigador.

	V1A	V1B	V1C	V2A	V2B	V2C	V2D	V2E	V2F	V2G	V2H	V2I	V2J	V2K	V6A	V6B	V7A	V7B	V7C	V8A	V8B	V8C	V10A
318	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	1	0	1	1	1
319	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	1	1	1
320	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	1	0	1	1	1	1
321	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	1	1	1
322	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	1	1	1
323	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	0	1	1	1	1
324	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1
325	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1	1	1	1
326	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1

Gráfico 41: Matriz de variables hijos procesados.
Elaborado por: Investigador.

3.2.3 Análisis de los datos y selección de características

Para la fase 3 de la metodología aplicada en el proyecto, haremos uso de las bondades que nos ofrece la técnica de la minería de datos como es, la visualización previa de los datos almacenados en los dos conjuntos de datos definidos en el dataset de variables padres y variables hijos, con la ayuda de la herramienta RStudio y el lenguaje de programación R que tiene un gran conjunto paquetes que permiten un análisis adecuado de datos. Utilizando los paquetes funModeling, tidyverse, Hmisc, procedemos a aplicar la función plot_num() sobre la matriz de las variables padres e hijos que previamente fueron almacenados en los dataset, los resultados se presenta en los gráficos 42 y 43.

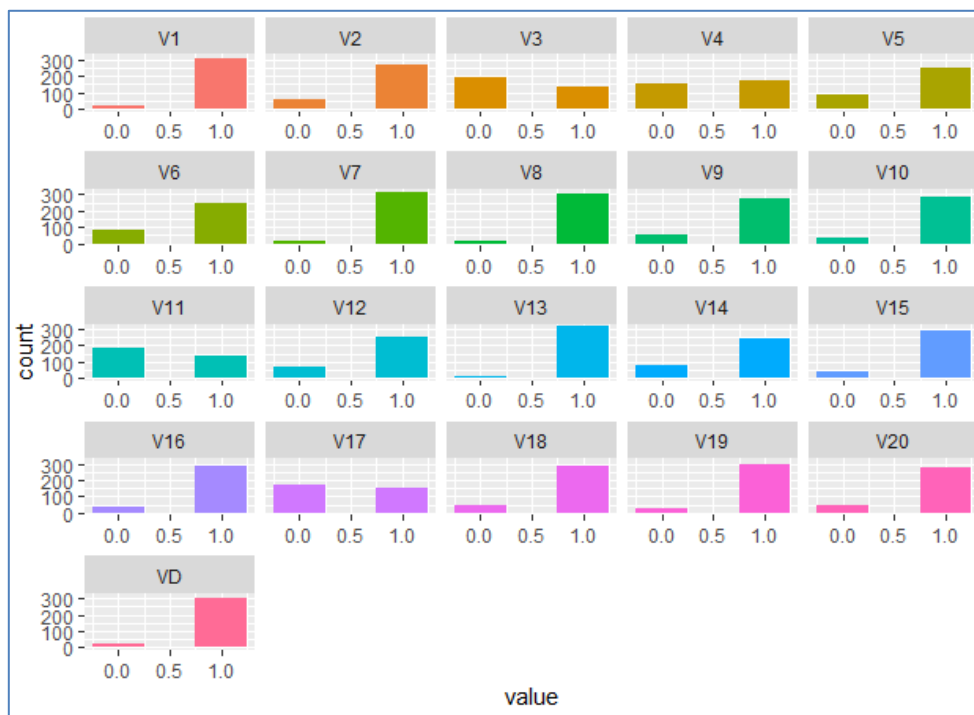


Gráfico 42: Visualización general de variables padres de deserción.
 Elaborado por: Investigador.



Gráfico 43: Visualización general de variables hijos de deserción.
 Elaborado por: Investigador.

Para determinar las variables más significativas en el proceso de investigación, se utiliza la herramienta Weka, donde con la ayuda del algoritmo GainRatioAttributeEval que evalúa el valor de un atributo midiendo la relación de ganancia con respecto a la clase y usando el método de búsqueda Ranker que clasifica los atributos según sus evaluaciones individuales, se procede a ingresar los dos Dataset (Padres - Hijos) de nuestra investigación a la herramienta antes indicada. Las Variables padres de deserción de mejor ranking e hijos, se presentan en las tablas 14 y 15.

Tabla 14: Ranking variables padres de deserción estudiantil.

Evaluator: weka.attributeSelection.GainRatioAttributeEval		
Search: weka.attributeSelection.Ranker -T - 1.7976931348623157E308 -N -1		
Instances: 326		
Attributes: 21 V1,V2,V3,V4,V5,V6,V7,V8,V9,V10,V11,V12,V13, V14,V15,V16,V17,V18,V19,V20,VD		
Evaluation mode: 10-fold cross-validation		
Attribute selection 10 fold cross-validation (stratified), seed: 1		
average merit	average rank	attribute
0.519 +- 0.035	1.3 +- 0.46	19 - V19
0.48 +- 0.027	1.7 +- 0.46	20 - V20
0.348 +- 0.044	3 +- 0	1 - V1
0.243 +- 0.017	4.5 +- 0.5	12 - V12
0.236 +- 0.019	4.6 +- 0.66	8 - V8
0.193 +- 0.021	6.1 +- 0.7	2 - V2
0.16 +- 0.022	8 +- 1.34	16 - V16
0.153 +- 0.017	8.2 +- 1.17	18 - V18
0.14 +- 0.021	9.5 +- 1.8	9 - V9
0.142 +- 0.026	9.7 +- 1.55	13 - V13
0.143 +- 0.012	9.7 +- 0.64	10 - V10
0.101 +- 0.018	12 +- 0.77	15 - V15
0.088 +- 0.011	13.1 +- 1.04	14 - V14
0.08 +- 0.005	14.1 +- 0.54	5 - V5
0.067 +- 0.018	14.5 +- 0.81	7 - V7
0.023 +- 0.006	16.1 +- 0.3	11 - V11
0.016 +- 0.004	16.9 +- 0.3	6 - V6
0.005 +- 0.002	18.2 +- 0.4	4 - V4
0.004 +- 0.003	18.9 +- 0.54	3 - V3
0.002 +- 0.001	19.9 +- 0.3	17 - V17

Elaborado por: Investigador.

Tabla 15: Ranking variables hijos de deserción estudiantil.

Evaluator: weka.attributeSelection.GainRatioAttributeEval		
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1		
Instances: 326		
Attributes: 70 V1A,V1B,V1C,V2A,V2B,V2C,V2D,V2E,V2F,V2G,V2H,V2I,V2J,V2K,V6A,V6B ,V7A,V7B,V7C,V8A,V8B,V8C,V10A,V10B,V10C,V10D,V11A,V11B,V11C,V11D,V11E,V12A, V12B,V12C,V13A,V13B,V13C,V14A,V14B,V18A,V18B,V18C,V18D,V18E,V18F,V18G,V18H, V18I,V19A,V19B,V19C,V19D,V19E,V19F,V19G,V19H,V19I,V19J,V19K,V19L,V19M,V19N,V 20A,V20B,V20C,V20D,V20E,V20F,V20G,VD		
Evaluation mode: 10-fold cross-validation		
Attribute selection 10 fold cross-validation (stratified), seed: 1		
average merit	average rank	attribute
0.303+-0.016	3.4+-1.43	64 - V20B
0.289+-0.021	4.9+-1.45	49 - V19A
0.282+-0.023	5.1+-0.83	68 - V20F
0.266+-0.034	5.8+-1.94	2 - V1B
0.273+-0.029	5.8+-1.54	53 - V19E
0.225+-0.014	9.3+-1	32 - V12A
0.219+-0.014	9.8+-1.25	9 - V2F
0.214+-0.032	10.9+-2.66	3 - V1C
0.217+-0.046	11+-3.77	4 - V2A
0.203+-0.018	11.4+-2.37	55 - V19G
0.199+-0.026	12.1+-2.21	66 - V20D
0.193+-0.014	12.7+-1	65 - V20C
0.164+-0.014	16.4+-1.56	8 - V2E
0.16+-0.01	17.2+-1.72	59 - V19K
0.155+-0.009	18.9+-2.21	60 - V19L
0.151+-0.012	19.2+-2.52	44 - V18E
0.155+-0.022	19.2+-4.87	24 - V10B
0.15+-0.014	19.6+-3.01	42 - V18C
0.143+-0.011	21.6+-2.62	40 - V18A
0.139+-0.018	22.2+-3.79	7 - V2D
0.132+-0.017	24.7+-3.85	62 - V19N
0.131+-0.015	24.8+-4.6	1 - V1A
0.131+-0.016	25.2+-3.71	54 - V19F
0.131+-0.015	25.3+-2.76	41 - V18B
0.128+-0.014	25.6+-3.85	58 - V19J
0.126+-0.012	26+-2.05	43 - V18D
0.124+-0.017	26.3+-4.67	23 - V10A
0.114+-0.013	29.2+-2.44	45 - V18F
0.098+-0.021	32.9+-5.28	35 - V13A
0.094+-0.01	33.9+-2.62	10 - V2G
0.094+-0.014	34.3+-3.72	61 - V19M
0.089+-0.006	35+-2.79	33 - V12B
0.09+-0.017	35.4+-4.43	50 - V19B
0.088+-0.009	35.9+-3.24	57 - V19I
0.083+-0.016	37.5+-4.34	52 - V19D
0.081+-0.01	37.7+-4.67	38 - V14A

Elaborado por: Investigador.

Tabla 15: Ranking variables hijos de deserción estudiantil. (Continuación)

average merit	average rank	attribute
0.079+-0.01	39+-2.97	21 - V8B
0.076+-0.009	39.9+-3.18	26 - V10D
0.075+-0.001	40+-1.67	34 - V12C
0.073+-0.01	41.2+-2.86	48 - V18I
0.07+-0.012	41.3+-3.61	6 - V2C
0.072+-0.007	41.4+-2.58	5 - V2B
0.067+-0.011	42.8+-3.09	47 - V18H
0.048+-0.008	46.9+-1.45	22 - V8C
0.043+-0.006	47.9+-1.58	67 - V20E
0.042+-0.012	48.3+-2.49	27 - V11A
0.042+-0.004	48.3+-1.27	56 - V19H
0.04+-0.008	49.1+-1.97	25 - V10C
0.036+-0.005	50.6+-1.5	29 - V11C
0.03+-0.004	52.9+-1.76	39 - V14B
0.029+-0.006	53.3+-1.73	28 - V11B
0.023+-0.005	55.5+-1.36	31 - V11E
0.02+-0.005	56.5+-2.38	14 - V2K
0.022+-0.007	56.6+-3.58	18 - V7B
0.02+-0.005	57.8+-2.75	15 - V6A
0.019+-0.007	58.2+-3.6	19 - V7C
0.018+-0.005	58.3+-2.53	12 - V2I
0.018+-0.004	59.2+-2.56	30 - V11D
0.018+-0.004	59.2+-2.4	16 - V6B
0.017+-0.004	59.3+-2.05	46 - V18G
0.01+-0.003	62.6+-1.28	13 - V2J
0.004+-0.002	64.5+-0.67	36 - V13B
0.003+-0.002	64.9+-0.94	37 - V13C
0.002+-0.001	66.3+-1.19	11 - V2H
0.002+-0.001	66.5+-1.02	51 - V19C
0+-0	67.9+-0.54	17 - V7A
0+-0	68.8+-0.4	20 - V8A

Elaborado por: Investigador.

3.2.4 Modelado

Puesto que el objetivo de la investigación es generar un modelo para la determinación de factores de deserción de la Universidad Técnica de Cotopaxi, se toma como variable predictiva la deserción estudiantil, cualitativa de tipo nominal.

3.2.4.1 Selección del Algoritmo para los modelos de redes neuronales

Para la selección del algoritmo, primero mencionaremos que la construcción de nuestro modelo estará basada en redes neuronales, mediante el algoritmo Perceptrón Multicapa y la función backpropagation [101]. La propagación hacia atrás de errores corresponde a un algoritmo de aprendizaje supervisado gráfico 44 y 45, esto quiere decir que las redes neuronales saben cuándo cometen un error y aprenden de estos errores, a este proceso se le llama backpropagation, el mismo que utilizaremos para realizar la determinación de factores de deserción estudiantil.

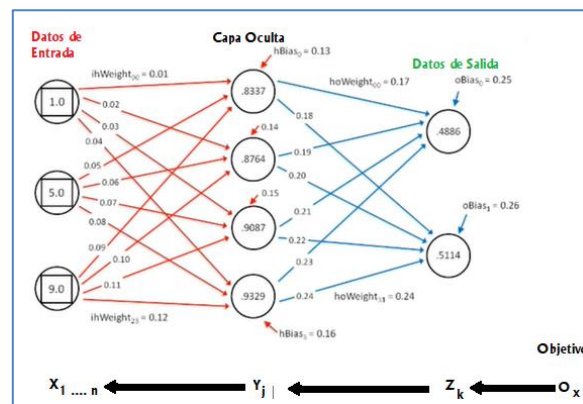


Gráfico 44: Esquema general de una neurona “backpropagation”.
Elaborado por: Investigador.

```

Algoritmo BACKPROPAGATION(red ejemplos,  $\eta$ ) {
  { $w_{ij}$ } ← INICIALIZAR;
  Mientras ¬ CONVERGENCIA(red) Hacer {
     $e$  ← SELECCIONAREJEMPLO(ejemplos);
    { $y_k$ } ← FORWARD( $e$ );
    { $d_k$ } ← DESEADAS( $e$ );
    Para cada  $n_k \in \text{CAPA}(\text{red}, k)$  Hacer {
       $\delta_k = (d_k - y_k) f'(net_k)$ ;
    }
    Para  $j = k - 1$  hasta 1 Hacer {
      Para  $n_j \in \text{CAPA}(\text{red}, j)$  Hacer {
         $\delta_j = f'(net_j) \sum_{j+1} \delta_{k+1} w_{j(j+1)}$ ;
      }
    }
    Para  $j = k$  hasta 1 Hacer {
       $w_{(j-1)j} = w_{(j-1)j} + \eta \delta_j y_{(j-1)}$ ;
    }
    red ← ACTUALIZARRED({ $w_{ij}$ });
  }
  Devolver red;
}
  
```

Gráfico 45: Algoritmo Backpropagation.
Elaborado por: Investigador.

Dada una red neuronal artificial y una función de error, el método calcula el gradiente de la función de error con respecto a los pesos de la red neuronal, esta es una regla para redes neuronales perceptrón multicapa. Es así como utilizaremos la función de activación (sigmoideal) que tiene una forma de “S”, como se muestra en el gráfico 46.

Para aprovechar la capacidad de aprender del algoritmo es recomendable utilizar las funciones no lineales en la neurona, de preferencia en la capa oculta. Por lo tanto, la función sigmoideal [102] será de gran ayuda como función de activación. Por ejemplo, si tenemos como dato de entrada el número 5 la neurona comparara con todos los números que existan de 0 a 5, para que la neurona tenga muchas opciones (todas las opciones posibles) y en diferentes lugares se activen diferentes puntos del sigmoide (ver gráficos 46 y 47). De esa manera las neuronas prueban cada uno de los datos y los compra con la base de datos de entrenamiento, con esa estructura final se transforma en un aprendizaje.

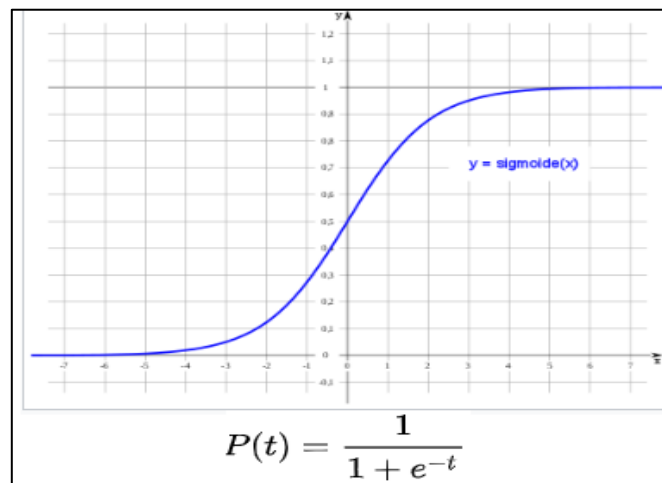


Gráfico 46: Función Sigmoide.

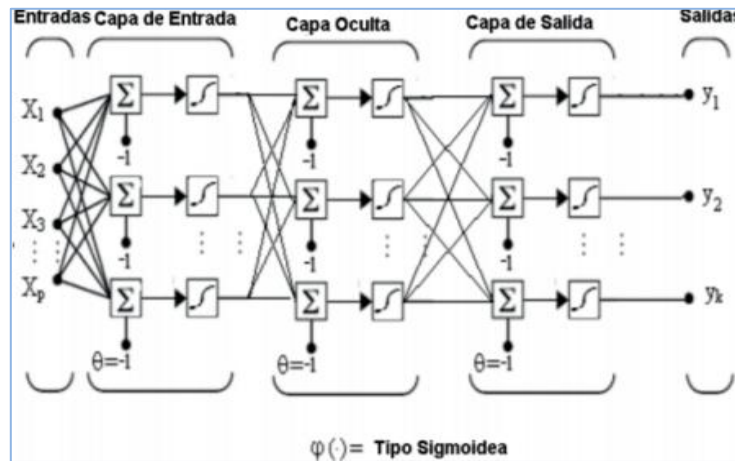


Gráfico 47: Estructura del perceptrón multicapa, utilizando como función de salida Sigmoide.

3.2.4.2 Generación del Modelo

Red neuronal perceptrón multicapa

En esta fase se hace el uso del algoritmo backpropagation, utilizando la herramienta RStudio con el lenguaje de programación R, en el cual se utilizaron las siguientes librerías neuralnet, MASS, ggplot2, DBI. Este algoritmo y librerías se utilizan basándose en las necesidades del proyecto, permitieron evaluar la información que posee dataset enfocado a los resultados que se quiere obtener y los objetivos que se busca cumplir.

El perceptrón multicapa se le puede diferenciar en dos fases, la de propagación calcula el resultado de salida de la red desde los valores de entrada hacia delante y la de aprendizaje que consiste en obtener los errores de salida del perceptrón, se va propagando hacia atrás (backpropagation) con el objetivo de modificar los pesos de cada conexión para que el valor estimado de la red se asemeje cada vez más al real.

Para la construcción de los modelos de redes neuronales tanto para los factores padre e hijos de deserción estudiantil, se presenta un patrón de entrada V : $v_1, \dots, v_{10}, \dots, v_{10a}, \dots, v_{19i}$, este se tramite a través de los pesos W_{ji} desde la capa de

entrada hacia la capa oculta. Las neuronas de esta capa intermedia transforman las señales recibidas mediante la aplicación de la función sigmoide de activación, de este modo se proporciona un valor de salida. Este se transmite a través de los pesos W_{kj} hacia la capa de salida, donde aplicando la misma operación que en el caso anterior, las neuronas de la última capa proporcional la salida de la red, código fuente de la construcción del modelo Anexo 2.

Separación de los datos en conjunto de aprendizaje y validación

Basándonos en la metodología de la (RNA), al fin de encontrar la red que tiene la mejor ejecución con casos nuevos es decir es capaz de “generalizar”, la muestra de los datos se tiende a subdividirla en dos grupos [103] entrenamiento y validación. Como se puede observar en las tablas 16 y 17, se presenta el procesamiento de los factores de deserción padres e hijos.

Tabla 16: Red neuronal – clasificación de la muestra - factores padres de deserción.

Resumen del procesamiento de los casos - Padres			
		N°	Porcentaje
Muestra	Entrenamiento	235	72,1%
	Prueba	91	27,9%
Válidos		326	100,0%
Excluidos		0	
Total		326	

Elaborado por: Investigador.

Tabla 17: Red neuronal – clasificación de la muestra - factores hijos de deserción

Resumen del procesamiento de los casos - Hijos			
		N°	Porcentaje
Muestra	Entrenamiento	227	69,6%
	Prueba	99	30,4%
Válidos		326	100,0%
Excluidos		0	
Total		326	

Elaborado por: Investigador.

Etapa de Aprendizaje

Para el aprendizaje de la red, los pesos son modificados de forma interactiva de acuerdo con los valores de grupo de entrenamiento, con el propósito de minimizar el error cometido entre la salida obtenida por la red. En la fase de aprendizaje suele ocurrir un problema “fenómeno del sobre ajuste”, es cuando el modelo se ajusta demasiado, esto quiere decir que el modelo aprende demasiado con los mismos datos de entrenamiento, perdiendo la habilidad de generalizar su aprendizaje a casos nuevos.

Para que el modelo de red neuronal no pierda la habilidad de predecir, es aconsejable utilizar un segundo conjunto de datos diferentes a los de entrenamiento, el grupo de validación, que permite controlar el proceso de aprendizaje. Durante la fase de aprendizaje la red va modificando los pesos en función a los datos de entrenamiento y de forma alternada se va obteniendo el error que comete la red ante los datos de validación.

Una vez aplicado el algoritmo backpropagation, a continuación, se proporciona los datos de entrada al modelo construido, los mismos que serán determinados mediante ensayo y error. La utilización del conjunto de datos de validación ayudará a conocer el valor más óptimo de cada uno de estos parámetros: valor de pesos iniciales, arquitectura de la red, valor de la tasa de aprendizaje y del momento y la función de activación de las neuronas de la capa oculta y de la de salida. De esta manera, la configuración de los parámetros que tenga el menor error ante el conjunto de datos de validación, serán seleccionados para pasar a la fase de pruebas, como se puede observar en las tablas 18 y 19.

Tabla 18: Red neuronal – resumen del modelo - factores padres de deserción

Resumen del modelo – Factores Padres		
Entrenamiento	Suma de errores cuadráticos	3,414
	Porcentaje de pronósticos incorrectos	,9%
	Regla de parada utilizada	1 pasos consecutivos sin disminución del error ^a
	Tiempo de entrenamiento	00:00:00,055
Prueba	Suma de errores cuadráticos	,619
	Porcentaje de pronósticos incorrectos	,0%

Elaborado por: Investigador.

Tabla 19: Red neuronal – resumen del modelo - factores hijos de deserción.

Resumen del modelo – Factores Hijos		
Entrenamiento	Error de entropía cruzada	4,954
	Porcentaje de pronósticos incorrectos	,0%
	Regla de parada utilizada	1 pasos consecutivos sin disminución del error ^a
	Tiempo de entrenamiento	00:00:00,032
Prueba	Error de entropía cruzada	11,487
	Porcentaje de pronósticos incorrectos	2,0%

Elaborado por: Investigador.

Rede neuronal Deep Learning

Como se expuso anteriormente Deep learning se caracteriza por emplear modelo de redes neuronales con muchas capas para así lograr un aprendizaje más profundo sobre el objetivo deseado. Para que este proceso funcione de manera adecuada es necesario tener un almacén de datos en nuestro caso los dos Dataset de factores de deserción padres y factores hijo que a partir de estos datos los algoritmos empleados pueden aprender de manera automática. Este modelo en particular requiere gran

capacidad computacional que en esta época no sería un problema, puesto que ya existen super computadoras que nos ayudara en el procesamiento de información.

Para nuestro proyecto hacemos uso de la poderosa librería de R como es keras que fue desarrollada para realizar experimentos rápidos y de alto nivel con redes neuronales, también es muy eficiente en la construcción de modelos de aprendizaje profundo. También aremos uso de otras librerías como son: neuralnet, mlbench, dplyr, magrittr, entre otras.

Separación de los datos en conjunto de aprendizaje y validación (Deep Learning)

Empleando las mejores prácticas de la metodología de la (RNA), con el fin de encontrar un resultado más eficiente se procede a dividir los datos en dos grupos: entrenamiento y pruebas, el primero se utiliza para obtener el modelo y el segundo para probarlo, como se puede observar en las tablas 20 y 21, correspondientes a los factores de deserción padres e hijos.

Tabla 20: Deep Learning – clasificación de la muestra - factores padres.

Data	
datos_padres	num [1:326, 1:21] 0.224 0.224 0.236 0.229 0.243 ...
entrenamiento_padres	num [1:233, 1:20] 0.224 0.236 0.229 0.236 0.243 ...
test_padres	num [1:93, 1:20] 0.224 0.229 0.243 0.243 0.224 ...

Resumen del procesamiento de los casos - Padres			
		N°	Porcentaje
Muestra	Entrenamiento	233	71.4%
	Prueba	93	28.6%
Válidos		326	100,0%
Excluidos		0	
Total		326	

Elaborado por: Investigador.

Tabla 21: Deep Learning – clasificación de la muestra - factores hijos.

Data	
datos_hijos	num [1:326, 1:70] 0.13 0.13 0.132 0.134 0.127 ...
entrenamiento	num [1:221, 1:69] 0.13 0.132 0.127 0.134 0.139 ...
test	num [1:105, 1:69] 0.13 0.134 0.144 0.149 0.131 ...

Resumen del procesamiento de los casos - Hijos			
		N°	Porcentaje
Muestra	Entrenamiento	221	67.8%
	Prueba	105	32.2%
Válidos		326	100,0%
Excluidos		0	
Total		326	

Elaborado por: Investigador.

Etapa de entrenamiento

Una vez separado los datos en un 70% y 30% de su total almacenado en los dos dataset, se procede a ejecutar el modelo con los datos de entrenamiento donde se puede observar una mejora incremental, para una predicción eficiente los pesos deben ser reiniciados aleatoriamente en el modelo, los mismos que afectaran a las relaciones entre las entradas y las salidas, los pesos se irán ajustando automáticamente por el algoritmo seleccionado, cuando más se entrena al modelo mejor será las predicciones. En los gráficos 48 y 49 se puede observar las épocas de entrenamiento del modelo de deserción estudiantil de los factores padre e hijos.

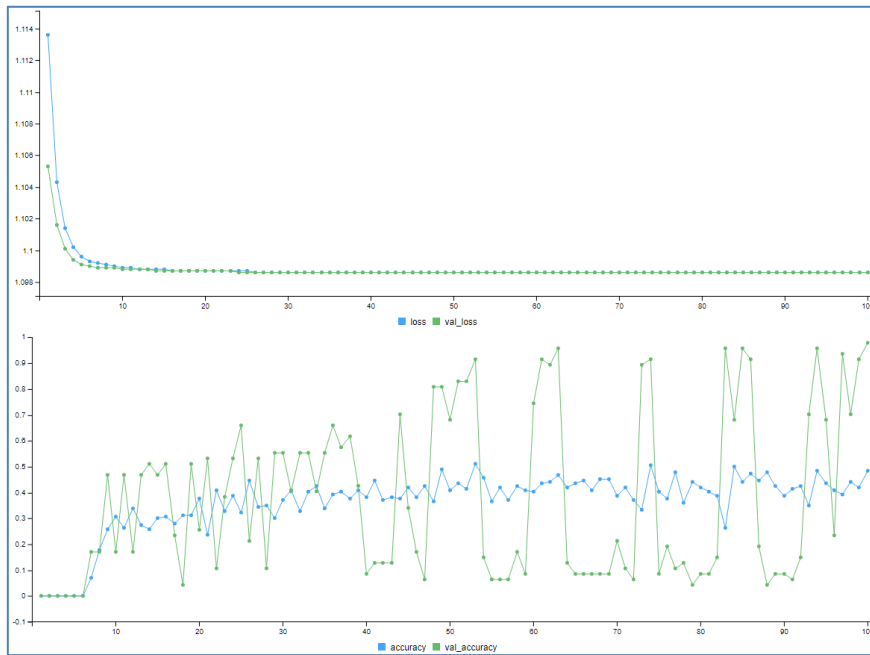


Gráfico 48: Deep learning, entrenamiento de factores padres de deserción.
 Elaborado por: Investigador.

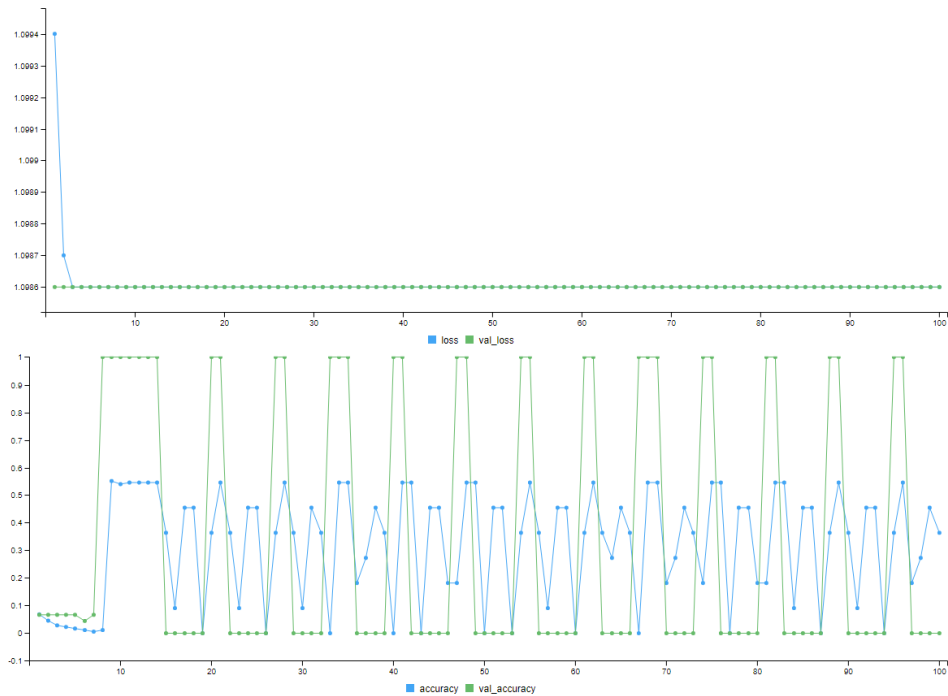


Gráfico 49: Deep learning, entrenamiento de factores hijos de deserción.
 Elaborado por: Investigador.

3.2.5 Evaluación

Modelo red neuronal perceptrón multicapa

Una vez construido el modelo de la red neuronal para la determinación de factores de deserción estudiantil en la Universidad Técnica de Cotopaxi, debemos evaluar la generalización de la red de una forma completamente objetiva a partir del tercer conjunto de datos independiente, el grupo de datos de test. Para la estimación del modelo se utilizará la media cuadrática del error para evaluar la ejecución del modelo, como se puede observar en las gráficas 50 y 51.

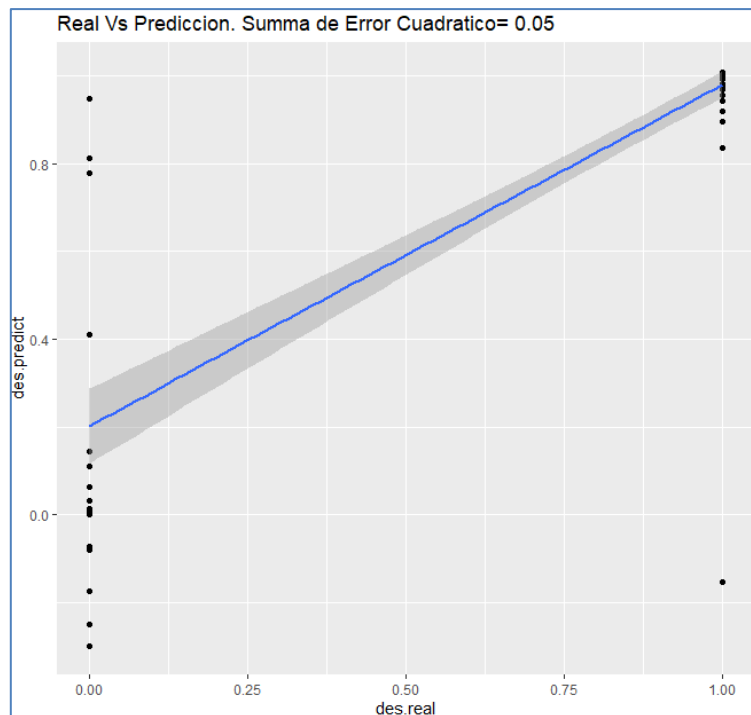


Gráfico 50: Error cuadrático –factores padres de deserción.
Elaborado por: Investigador.

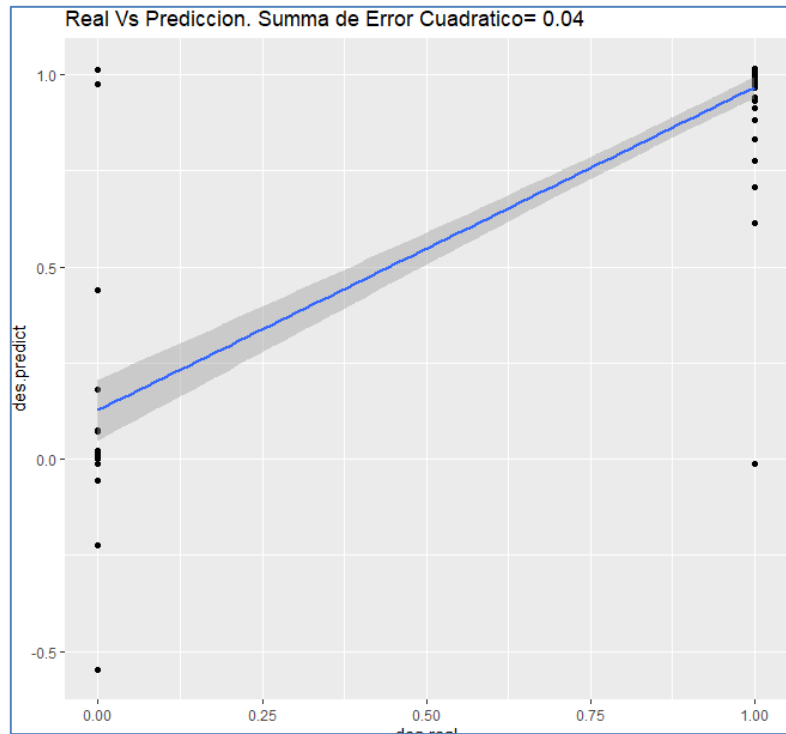


Gráfico 51: Error cuadrático – factores hijos de deserción.
 Elaborado por: Investigador.

Modelo Red neuronal Deep Learning

Para la evaluación del modelo de aprendizaje profundo hay que tomar en cuenta la configuración de la red como: el número niveles, el número de neuronas en cada nivel y por último las épocas con que se entrena el modelo, como se puede observar en el gráfico 51 los niveles de exactitud para el modelo de factores padres con un 0.957 esto quiere decir que tiene una tasa de predicción de un 95,7%, de igual forma para el modelo de factores hijos gráfico 52 se puede observar que tiene una exactitud de 0.978 equivalente a una tasa de predicción de un 97,8%.

```
> # Evaluar modelo con prueba - datos_padres
> model1 <- model %>%
+   evaluate(test_padres, etiqueta_test_p)
3/3 [=====] - 0s 613us/step - loss: 1.0986 - accuracy: 0.9570
```

Gráfico 52: Nivel de exactitud, factores padres de deserción.
 Elaborado por: Investigador.

```

> # Evaluar modelo de prueba con datos_hijos
> model2 <- model %>%
+ evaluate(test, etiqueta_test)
3/3 [=====] - 0s 1ms/step - loss: 1.0986 - accuracy: 0.9785

```

Gráfico 53: Nivel de exactitud, factores hijos de deserción.
 Elaborado por: Investigador.

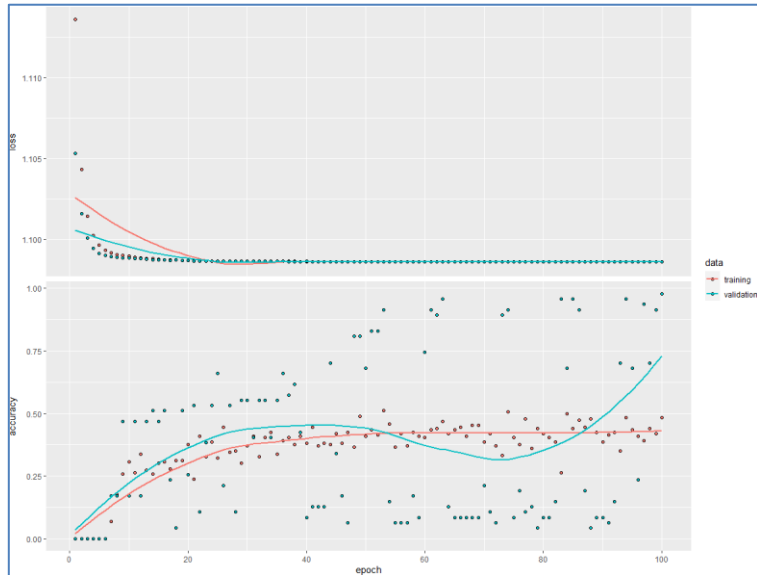


Gráfico 54: Evaluación del modelo deep learning, factores padres de deserción.
 Elaborado por: Investigador.

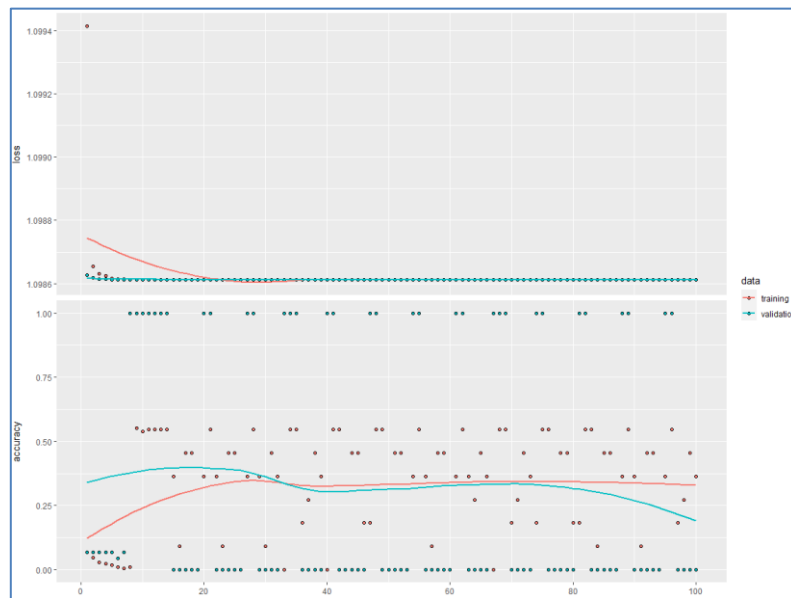


Gráfico 55: Evaluación del modelo deep learning, factores hijos de deserción.
 Elaborado por: Investigador.

3.2.5.1 Cross-Validation del modelo de deserción estudiantil.

La técnica cross-validation ayuda a profundizar aún más la precisión del modelo, para eso se procede a separar los datos de entrenamiento 70% y de pruebas 30%, de los cuales creamos otros 10 subconjuntos datos, a continuación creamos una función (ver Anexo 3) para ejecutar sobre cada subconjunto de datos donde devuelve el *Accuracy* y *Kappa* del modelo, como se muestra en la tabla 22.

Tabla 22: Cross-validation – factores padres e hijos.

10 Subconjunto de datos para el modelo padres e hijos										
Factores de deserción padres										
	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10
Accuracy	1	0.931	1	1	1	1	1	0.965	1	1
Kappa	1	0.632	1	1	1	1	1	0.868	1	1
Factores de deserción hijos										
	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10
Accuracy	1	0.931	1	1	0.965	1	1	1	0.933	1
Kappa	1	0.632	1	1	0.781	1	1	1	0.714	1

Elaborado por: Investigador.

Tabla 23: Precisión del modelo para los factores padre e hijos.

Accuracy factores padres	
Accuracy	Kappa
0.9896552	0.9901690
Accuracy factores hijos	
Accuracy	Kappa
0.9829885	0.9829152

Elaborado por: Investigador.

Al analizar a tabla 23 se termina que el clasificador ha logrado una exactitud para los factores padres de un 0.989, esto quiere decir que se ha obtenido una precisión del 98,9 % en la predicción del modelo de deserción estudiantil, así como también para los factores hijos se tiene una exactitud del 0.982 este valor representa una precisión del 98.2 % para la predicción del modelo. Por lo tanto, podremos usar con toda confianza este modelo para predecir los factores de deserción estudiantil universitaria.

3.2.5.2 Especificación del modelo

redes neuronales se puede observar los Gráficos 56 y 57. Para modelar la red neuronal se consideró variables binarias, es decir 1 para predecir que los estudiantes de la universidad no abandonarían la institución y 0 para predecir que los estudiantes si abandonarían la universidad.

La especificación del modelo

```
*Multilayer Perceptron Network.
MLP VD BY V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20
/RESCALE DEPENDENT=STANDARDIZED
/PARTITION TRAINING=7 TESTING=3 HOLDOUT=0
/ARCHITECTURE AUTOMATIC=NO HIDDENLAYERS=1 (NUMUNITS=AUTO) HIDDENFUNCTION=SIGMOID OUTPUTFUNCTION=IDENTITY
/CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0
INTERVALOFFSET=0.5 MEMSIZE=1000
/PRINT CPS NETWORKINFO SUMMARY SOLUTION IMPORTANCE
/PLOT NETWORK PREDICTED RESIDUAL
/STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=
0.0010
/MISSING USERMISSING=EXCLUDE .
```

Gráfico 56: Modelo de deserción – factores padres de deserción.
Elaborado por: Investigador.

```
*Multilayer Perceptron Network.
MLP VD BY V1A V1B V1C V2A V2B V2C V2D V2E V2F V2G V2H V2I V2J V2K V6A V6B V7A V7B V7C V8A V8B V8C V10A V10B
V10C V10D V11A V11B V11C V11D V11E V12A V12B V12C V13A V13B V13C V14A V14B V18A V18B V18C V18D V18E V18F V18
G V18H V18I V19A V19B V19C V19D V19E
V19F V19G V19H V19I V19J V19K V19L V19M V19N V20A V20B V20C V20D V20E V20F V20G
/PARTITION TRAINING=7 TESTING=3 HOLDOUT=0
/ARCHITECTURE AUTOMATIC=YES (MINUNITS=1 MAXUNITS=50)
/CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVA
LCENTER=0 INTERVALOFFSET=0.5 MEMSIZE=1000
/PRINT CPS NETWORKINFO SUMMARY SOLUTION IMPORTANCE
/PLOT NETWORK PREDICTED RESIDUAL
/STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 E
RRORRATIO=0.0010
/MISSING USERMISSING=EXCLUDE .
```

Gráfico 57: Modelo de deserción – factores hijos de deserción.
Elaborado por: Investigador.

Para la construcción del modelo de predicción se realiza la partición del conjunto de datos en un 70% (227 casos) determinados para el entrenamiento, el 30% (99 casos) seleccionados para las pruebas. En la arquitectura de la red neuronal se utilizaron los datos tipificados, para el proceso de entrenamiento de la red neuronal, se utilizó la función sigmoial, así como también se consideró una primera capa oculta con 5 neuronas y una segunda capa oculta con 3 neuronas, lo que indica que las iteraciones

se detendrán cuando el cambio del error sea menor a 5%. Entre una iteración de optimización y otra, también se utiliza el algoritmo backpropagation, que actualiza los pesos considerando únicamente el signo del cambio, es decir, si el cambio del error es en aumento (+) o disminución (-) entre una iteración y otra. Teniendo como resultados la arquitectura del modelo de red neuronal para los factores padres e hijos de deserción estudiantil universitaria, como se puede observar en las gráficas 58 y 59.

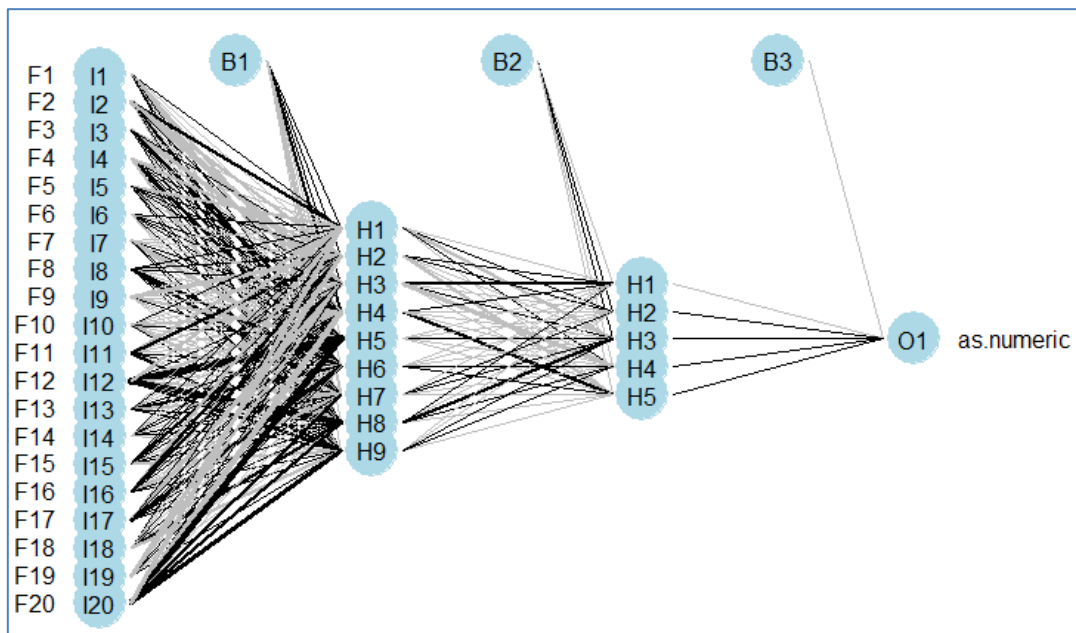


Gráfico 58: Red neuronal - factores padres de deserción.
Elaborado por: Investigador.

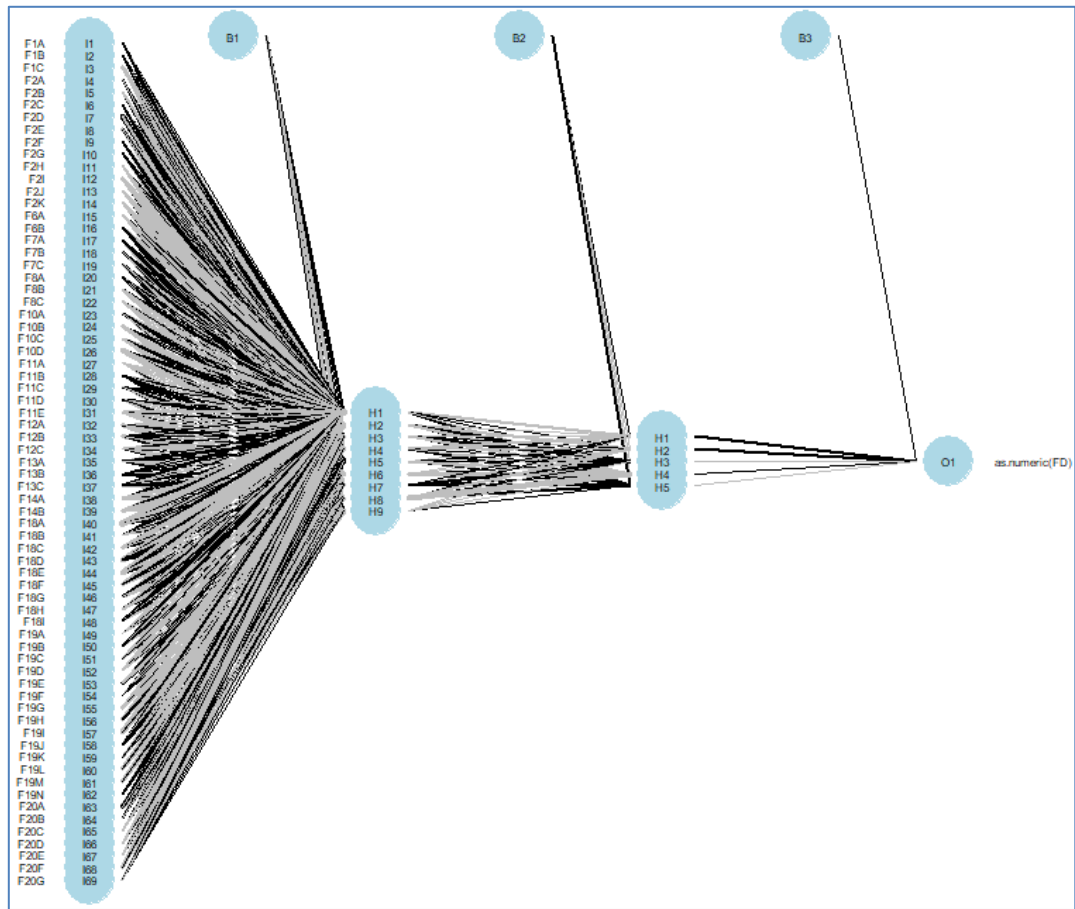


Gráfico 59: Red neuronal - factores hijos de deserción.
 Elaborado por: Investigador.

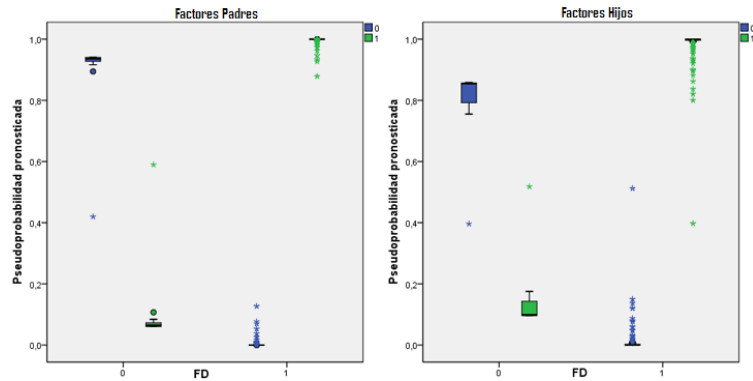


Gráfico 60: Pseudoprobabilidad predicha de los factores padres e hijos
 Elaborado por: Investigador.

Factores padres		
Red Nueronal	Entrenamiento	Prueba
Perceptrón muticapa (backpropagation)	97,2%	98,6%
Factores hijos		
Perceptrón muticapa (backpropagation)	96,7%	98,4%

Gráfico 61: Precisión de la predicción de los modelos de predicción
 Elaborado por: Investigador.

Red neuronal Deep Learning

La especificación del cómo está construido el modelo de redes neuronales Deep learning, para la determinación de factores de deserción estudiantil en la Universidad Técnica de Cotopaxi gráfico 62 y 63, donde se puede observar que en la primera capa está compuesta de 50 neuronas, en la segunda capa de 10, en la tercera de 5 y en una cuarta y última capa de 3, haciendo uso del todo el poder del aprendizaje profundo.

```
> # Creando el modelo secuencial de deserción estudiantil
> model <- keras_model_sequential()
> model %>%
+ layer_dense(units=50, activation = 'relu', input_shape = c(20)) %>%
+ layer_dense(units=10, activation = 'relu') %>%
+ layer_dense(units = 5, activation = "relu") %>%
+ layer_dense(units = 3, activation = "sigmoid")
> summary(model)
Model: "sequential_7"
Layer (type) Output Shape Param #
-----
dense_25 (Dense) (None, 50) 1050
dense_26 (Dense) (None, 10) 510
dense_27 (Dense) (None, 5) 55
dense_28 (Dense) (None, 3) 18
-----
Total params: 1,633
Trainable params: 1,633
Non-trainable params: 0
```

Gráfico 62: Modelo deep learning, factores padres de deserción.
 Elaborado por: Investigador.

```
> # Creando el modelo secuencial de deserción estudiantil
> model <- keras_model_sequential()
> model %>%
+ layer_dense(units=50, activation = 'relu', input_shape = c(89)) %>%
+ layer_dense(units=10, activation = 'relu') %>%
+ layer_dense(units = 5, activation = "relu") %>%
+ layer_dense(units = 3, activation = "sigmoid")
> summary(model)
Model: "sequential_9"
Layer (type) Output Shape Param #
-----
dense_33 (Dense) (None, 50) 3500
dense_34 (Dense) (None, 10) 510
dense_35 (Dense) (None, 5) 55
dense_36 (Dense) (None, 3) 18
-----
Total params: 4,083
Trainable params: 4,083
Non-trainable params: 0
```

Gráfico 63: Modelo deep learning, factores hijos de deserción.
 Elaborado por: Investigador.

3.3 Conclusiones del Capítulo III

- La metodología para desarrollar un proyecto de minería de datos, permitió valorar el ciclo de vida de este trabajo de investigación, de forma que se obtiene los datos a través de procedimiento experimental mediante el uso de técnicas de minería de datos e inteligencia artificial como son las redes neuronales dar cumplimiento a los objetivos planteados.
- Con base en los resultados obtenidos en el proyecto de investigación se puede concluir que el modelo de red neuronal desarrollado bajo el algoritmo Perceptrón Multicapa mediante la función backpropagación arrojó una tasa de precisión total de la predicción de la deserción del 98.42%, lo que indica que el modelo propuesto es robusto en cuanto a la adaptación, ya que este logra actualizar los pesos considerando únicamente signo de cambio del error, esto quiere decir que la red sabe cuándo cometió un error y retorna hacia atrás para modificar los pesos y así obtener una salida más representativa.
- El análisis de los resultados de la red neuronal perceptrón multicapa con una tasa de predicción de un 98,6% de los factores padres y un 98.4% de los factores hijos. Indican que el modelo desplegado es fiable y puede ser utilizado como una herramienta de apoyo a la toma de decisiones de las autoridades de Universidad Técnica de Cotopaxi.

CONCLUSIONES GENERALES:

- La revisión de la literatura permitió determinar los referentes teóricos, basadas en 106 fuentes de información científica, que ha permitido sustentar el problema de la investigación y la hipótesis planteada en el proyecto propuesto.
- A través de la aplicación de encuestas dirigida a expertos en temas de deserción universitaria, se propone un modelo conceptual de deserción que consta de 89 factores generales, de los cuales 20 factores corresponde a factores padres y el resto a factores hijos. El proceso de la regresión lineal a través del método de mínimos cuadrados ordinarios permitió determinar la influencia de los factores en la deserción, como resultado de este proceso se obtiene 31 factores con nivel de causalidad positiva y negativa en la variable dependiente - deserción universitaria.
- Se valida el modelo conceptual de deserción a través de la implementación de minería de datos, como resultado del proceso se obtiene una tasa de precisión para predicción de la deserción universitaria del 99,6 %, utilizando la técnica de red neuronal – deep learning los resultados permiten determinar que el modelo propuesto es óptimo en términos de confiabilidad.

RECOMENDACIONES:

- El desarrollo de un módulo de tableros de control estudiantiles en el sistema académico universitario, que permita predecir con anterioridad a los estudiantes con riesgo de desertar de la institución.
- Se propone realizar un estudio de investigación sobre el rumbo que toman los alumnos que abandonan la institución, a fin de conocer si estos estudiantes finalizan sus estudios en otra institución de nivel superior.
- La comprobación del modelo conceptual aplicando otros entornos de inteligencia artificial, con por ejemplo Árbol de Decisión, Naive Bayes, Clustering, entre otros.

BIBLIOGRAFÍA

- [1] E. Castaño, S. Gallón, K. Gómez, and J. Vásquez, "Deserción estudiantil universitaria: Una aplicación de modelos de duración," 2004.
- [2] L. Patiño and A. Cardona, "Revisión de algunos estudios sobre la deserción estudiantil Universitaria en Colombia y Latinoamérica," *Grupo de Investigación GESE*, vol. 21, Sep. 2012.
- [3] N. Torres, "El fracaso escolar: entre preocupaciones políticas y esfuerzos investigativos," *Revista de Filosofía y Letras*, p. 27, June 2017.
- [4] G. Páramo and C. Correa, "Deserción Estudiantil Universitaria. Conceptualización," *Escuela de Ingeniería - Universidad EAFIT*.
- [5] A. Seidman, "College student retention: A formula for success," 2012.
- [6] F. Cevallos, "Deserción, calidad y reforma universitaria. Apuntes para el debate," in *Cuadernos del Contrato Social por la Educació*. Quito, Ecuador, 2014, ch. Número 10.
- [7] J. Vivas, "El abandono de estudiantes universitarios," 2005.
- [8] U Giraldo, D Abad, and E Díaz,. Bogotá: CNA, 2005.
- [9] B.R. Clark, , *Organizational Pathways of Transformation*, Ed. Great Britan, 1998.
- [10] UPN and Universidad Pedagógica Nacional (2006),. Bogotá: UPN.
- [11] K. A. y Reason, R. D. Renn, *College Students in the United States.*, 2013.
- [12] V. Tinto, "Research and practice of student retention:What next?," , " *COLLEGE STUDENT RETENTION*, vol. Vol. 8(1), 2007.
- [13] R. Zárate and E. Mantilla, "La deserción estudiantil UIS, una mirada desde la responsabilidad social universitaria.," *Revista del Instituto de Estudios en Educación Universidad del Norte*, no. 21, pp. pp.121-134, julio-diciembre 2014.
- [14] T. Achilie, "Abandono de los estudios universitarios," Mayo 2016.
- [15] Ma. G. E. Huesca and Ma. B. Castaño, "Causas de Deserción de Alumnos de

Primeros Semestres de una Universidad Privada," *SCIELO*, 2007.

- [16] A. Garzón and J. Gil, "El papel de la procrastinación académica como factor de la deserción universitaria," *Revista Complutense de Educación*, vol. ISSN: 1130-2496, <https://revistas.ucm.es/index.php/RCED/article/view/49682>, 2017.
- [17] D. A. Viteri and M. A. Uquillas, "Estudio sobre la deserción estudiantil en Pontificia Universidad Católica del Ecuador - Matriz, en los niveles 1ro, 2do y 3ero de todas las Facultades y Escuelas del primer semestre del año académico 2007 - 2008," Quito, 2011.
- [18] V. Tinto, "A Theoretical Synthesis of Recent Research," 1975.
- [19] Asamblea Nacional, "*Constitución del Ecuador*", art. 356., 2008.
- [20] Gobierno del Ecuador, "*Ley orgánica de Educación Superior*", *Registro oficial.*, 2010.
- [21] El Telégrafo. (2016, Nov.) La deserción universitaria tiene cuatro causas.
- [22] A. A. Guadagni, "'Ingreso a la Universidad en Ecuador, Cuba y Argentina'," *CEEA:Argentina*, 2016.
- [23] J. S. Taylor, M. L. Machado, and M. W. Peterson, "Leadership and strategic management: Keys to institutional priorities and planning.," *European Journal of Education*, pp. 43 (3), 369-386, 2008.
- [24] M. Schneider, "The Cost of First-Year Student Attrition in America's Four-Year Colleges and Universities," no. https://www.air.org/sites/default/files/AIR_Schneider_Finishing_the_First_Lap_Oct101.pdf, 2010.
- [25] S. Donoso, G. Donoso, and O. Arias, "Iniciativas de retención de estudiantes de educación superior," *ResearchGate*, 2010.
- [26] C. Márquez, C.I Romero, and S. Ventura, "Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos," *IEEE-RITA*, vol. Vol. 7, no. Núm. 3, Nov. 2012.
- [27] S. Formia, L. Lanzarini, and W. Hasperué, "Caracterización de la deserción universitaria en la UNRN utilizando Minería de Datos. Un caso de estudio.," *Revista Iberoamericana De Tecnología En Educación Y Educación En Tecnología*, vol. 11, no. Recuperado a partir de <https://teyet-revista.info.unlp.edu.ar/TEyET/article/view/297>, pp. 92-98, 2011.

- [28] M. Jadrić, M. Željko, and M. Čukušić, "Student Dropout Analysis with Application of Data Mining Methods," *Preliminary communication,UDC 65.012.34:378*, 2010.
- [29] K. B. Eckert and R. Suénaga, "Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería," *Universidad Gastón Dachary, Departamento de Ingeniería y Ciencias de la Producción*, vol. 8, no. doi: 10.4067/S0718-50062015000500002, marzo 2015.
- [30] J. Creswell. (2008, febrero) University of Michigan. [Online]. <https://it.umich.edu>
- [31] T Christ, "A Recursive Approach to Mixed Methods Research in a Longitudinal Study of Postsecondary Education Disability Support Services," *Journal of Mixed Methods Research*, 2007.
- [32] R. Cameron, "El uso de métodos mixtos en la investigación en educación vocacional]," *Recuperado de <http://www.avetra.org.au/papers-2009/papers/12.00.pdf>*, 2009.
- [33] Z. Pereira, "Los diseños de método mixto en la investigación en educación: Una experiencia concreta," *Centro de Investigación y Docencia en Educación, Universidad Nacional Heredia, Costa Rica*, 2011.
- [34] J. M. Moine, S. Gordillo, and A. S. Haedo, "Estudio comparativo de metodologías para minería de datos," *Grupo de investigación en Minería de Datos, UTN Rosario Facultad de Ciencias Exactas*, 2016.
- [35] L. C. Paredes, "Estudio de la causa que genera la deserción académica de los estudiantes de la Escuela de Psicología, Casao PUCESA," Ambato, 2017.
- [36] A. Corena, P. Losada, and E. Mahecha, "Definiciones fundamentales sobre deserción estudiantil.," *Fundación Universidad Autónoma de Colombia, Bogotá.*, 2012.
- [37] P. A. Chinome, C. Ruiz, and L. Fernández, "Priorización de variables en el diseño de un sistema de gestión integral de la deserción estudiantil," *Educación en Ingeniería*, vol. 11 (22), pp. 69-77, mayo 2016.
- [38] V. Tinto, "Definir la Deserción: Una Cuestión de Perspectiva.," *Revista de educación superior*, 1989.
- [39] A. Astin, "Preventing students from dropping out," in *Jos-sey-Bass*. San

Francisco, 1975.

- [40] I. C. Montes, P. M. Almonacid, S. Gómez, F.. Zuluaga, and E. Tamayo, "Análisis de la deserción estudiantil en los programa de pregrado de la universidad EAFIT," *Serie Cuadernos de Investigación*, no. ISSN 1692-0694, 2010.
- [41] P. León, J. Romero , and I. Hernández, "'¿Desertores o decepcionados? Distintas causas para abandonar los estudios universitarios'," *Revista de la educación superior*, vol. 40, no. 160, pp. pp. 29-49, 2011.
- [42] W. Fontalvo, M. P. Castillo, and S. Polo, "Análisis comparativo entre las características más relevantes de deserción estudiantil en el programa de Ingeniería Industrial de la Universidad Autónoma del Caribe. Estudiantes activos en el periodo 2013-01 y desertores académicos de los periodos 2011-01," *Escenarios*, vol. 12, no. 1, junio 2014.
- [43] M. Rojas and D. C. González, "Deserción estudiantil en la Universidad de Ibagué, Colombia: una lectura histórica en perspectiva cuantitativa," 2011.
- [44] D. S. Sevilla, V. A. Puerta, and J. Dávila, "Influencia de los factores socioeconómicos en la deserción estudiantil de la carrera de ciencias sociales," *CIENCIA E INTERCULTURALIDAD*, vol. 6, no. 1, enero 2010.
- [45] G. Sanchez, W. Navarro, and A. D. García, "Factores de desercion estudiantil en la universidad surcolombiana," 2009.
- [46] Centro de Microdatos, "Estudio sobre causas de la deserción universitaria," 2008.
- [47] A. Casillas , S. Robbins, J Allen, and C Schmeiser, "Predicting early academic failure in high school from prior academic achievement psychosocial characteristics, and behavior," *Journal of Educational Psychology*. doi: 10.1037/a0027180, 2012.
- [48] J.M Braxton, R. M. Johnson, and A. Shaw-Sullivan, "Appraising Tinto's theory of colleger student departure," *Higher Education Handbook of theory and research*, vol. Vol.12, 1997.
- [49] E. R. Balmori, M. T. Carza, and E. Reyes, "El modelo de deserción de tinto como base para la planeación institucional: El caso de dos instituciones de educación superior tecnológica.," *XI Congreso Nacional de Investigación Educativa*, 2016.

- [50] E. Himmel, "Modelos de análisis de la deserción estudiantil en la educación superior.",
- [51] M. Fishbein and I. Ajzen,. EE.UU: Addison-Wesley, Reading, 1975.
- [52] C. A. Ethington , "A psychological model of student persistence," *Reaseach in Higher Education*, 1990.
- [53] J. Eccles , T. F. Adler, R. Futerrman , S. B. Goff, and C. Midgley,. San Francisco, EE.UU, 1983.
- [54] G. Matos, R. Chalmeta, and O. Coltell, "Metodología para la Extracción del Conocimiento Empresarial a partir de los Datos," vol. 17, no. 2, 2006.
- [55] V. Valcárcel, "Data Mining y el descubrimiento del conocimiento," *Industrial Dat*, 2004.
- [56] P. V. Britos and R. Garcia,. Buenos Aires, Argentina: Nueva Librería, 2005.
- [57] H. O. Nigro, D. Xodo, G. Corti, and D. Terren, "KDD (Knowledge Discovery in Databases): Un proceso centrado en el usuario," in *UNICEN – Tandil.*: Departamento de Computación y Sistemas, 2015.
- [58] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The complete guide to dimensional modeling.*: 2º ed. Wiley Publishing, 2002.
- [59] S. Valero , A. Salvador , and M. García , "Minería de datos: predicción de la deserción escolar," in *Universidad Tecnológica de Izúcar de Matamoros mediante el.*: 8, 2008.
- [60] kdnuggets. (2014, Oct.) www.kdnuggets.com. [Online]. <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>
- [61] J. Ponce and A. Karahoca,. Vienna: I-Tech, 2009, p. pág. 438.
- [62] Omar Salvador Gómez Gómez, "Propuesta de mejora sobre la primera etapa del modelo de proceso KDDM, CRISP-DM,".
- [63] J. C. Riquelme, R. Ruiz, and K. Gilbe, "Minería de Datos: Conceptos y Tendencias," *Revista Iberoamericana de Inteligencia Artificial.*, vol. 10, no. 29, 2006.
- [64] J. C. Giraldo and J. A. Jiménez, "Caracterización del Proceso de Obtención de Conocimiento y Algunas Metodologías para Crear Proyectos de Minería de

- Datos.," *Revista Latinoamericana de Ingeniería de Software*, vol. 1, no. 2, 2013.
- [65] W. Santamaria. (2015) Recuperado el Enero de 2016, deresearchgate.net. [Online]. researchgate.net
- [66] O. Maimon and L. Rokach, "Data Mining and Knowledge Discovery Handbook, 2nd ed," 2010.
- [67] J. L. Rivero, "Técnicas de aprendizaje automático para la detección de," *Revista Cubana de Ciencias Informáticas*, pp. 52-73, 2014.
- [68] Z. Ghahramani, "2004," in *Advanced Lectures on Machine Learning*. Springer-Verlag., 2004.
- [69] M. Bramer, *Principles of Data Mining*. London: SpringerVerlag.: Second Edition, 2013.
- [70] J. A. Hartigan and M. A. Wong,.: *Journal of the Royal Statistical Society*, 1979, pp. 100-108.
- [71] X. Basogain,.: *Escuela Superior de Ingeniería de Bilbao*, EHU, 1998.
- [72] R. Camana, "Potenciales Aplicaciones de la Minería de Datos en Ecuador," *Revista Tecnológica ESPOL*, vol. 29, no. 1, julio 2016.
- [73] E. Muñoz and C. Seijas, "Análisis comparativo de ajuste en entrenamiento de redes neuronales artificiales a partir de las librerías open NN y ALGLIB," *Revista de Ciencias de la Vida*, vol. 21, no. 1, pp. 49-60, 2015.
- [74] H. Paz and D. Jimenez , "Desarrollo de un sistema inteligente para la clasificación de documentos ya digitalizados aplicando redes neuronales supervisadas.," *Revista Tecnológica - ESPOL*, vol. 28, no. 1, 2015.
- [75] Instituto Politécnico Nacional. (2012, noviembre) *Simuladores de Redes Neuronales*.
- [76] R. Flores and J. M. Fernández, "Las Redes Neuronales Artificiales," *Fundamnetos y aplicaciones practicas*, Dec. 2015.
- [77] H., El-Dahshan, E.-S. A., El-Horbaty, E.-S. M., & Salem, A.-B. M. Mohsen, "Clasificación mediante redes neuronales de aprendizaje profundo para tumores cerebrales.," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68-71, 2018.

- [78] M. Nielsen, *Neural Networks and Deep Learning*, The original online book can be found at <http://neuralnetworksanddeeplearning.com/> ed., 2015.
- [79] J. Schmidhuber, "Aprendizaje profundo en redes neuronales: Una visión general," *Neural Networks*, vol. 61, no. 85, p. 117, 2015.
- [80] G. Rodríguez, *Metodología de la investigación cualitativa. Capítulo I*. Granada (España): Ediciones Aljibe, 1996.
- [81] R Hernández,. México D.F. México: Mc – Graw- Hill, 1991.
- [82] N. Cohen and G. Gómez, "Metodología de la investigación, ¿Para qué?," in *Metodología de la investigación, ¿Para qué?*, ISBN: 9789877231908, Ed. Buenos Aires, Argentina: Editorial Teseo, 2019.
- [83] R. Montero, "Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada," *Universidad de Granada*, 2016.
- [84] R. M. González and V. González, "Diagnóstico de necesidades y estrategias de formación docente en las universidades.," *Revista Iberoamericana de Educación (ISSN: 1681-5653)*, vol. 43/6, 2007.
- [85] E. Durán and D. Arias, "Orientación a las metas académicas, persistencia y rendimiento en estudiantes del Ciclo de Iniciación Universitaria.," 189-205 *ISSN: 1887-4592*, vol. Vol. 13 (2), Mayo 2015.
- [86] M. Albarracín and L. Muñoz, "Factores asociados al consumo de alcohol en estudiantes de los dos primeros años de carrera universitaria," *LIBERABIT*, julio 2008.
- [87] Y. Muñoz , "La ludopatía: revisión y análisis hacia un modelo integral.," *Drugs and Addictive Behavior*, vol. 1, no. 2, 2016.
- [88] Y. C. Chura, "Universidad nacional del altiplano de puno facultad de trabajo social escuela profesional de trabajo social el uso de los video juegos y su influencia en el ausentismo académico de los estudiantes de la facultad de medicina veterinaria y zootecnia de la UNA-P," Universidad Nacional del Antiplano, PUNO –PERÚ, 2019.
- [89] P. N. Maya, J. R. Aguilar, R. A. Zamora, and J. M. Barron, "Diseño de un Modelo predictivo aplicando Minería de Datos para identificar causas de Deserción Estudiantil Universitaria," *ECHNOLOGY & SOCIETY*, 2018.
- [90] M. S. Albán, "Contribuciones a la predicción de la deserción," Lima, Perú,

Tesis Doctoral 2019.

- [91] S. R. Arancibia and G. H. Rodríguez, "Representaciones sociales en torno a equidad, acceso y adaptación en educación Universitaria," *PSICOPERSPECTIVA*, vol. VOL. 12, no. 1, 2013.
- [92] T. Larroucau, "Estudio de los factores determinantes de la deserción en el sistema universitario chileno.," *Est. de Políticas Públicas*, vol. 1, no. 1, pp. 1-23, junio 2015.
- [93] L. F. Londoño, "Factores de riesgo presentes en la deserción estudiantil en la Corporación Universitaria Lasallista," *Revista Virtual Universidad Católica del Norte*, mayo 2013.
- [94] G. Devincenzi, G. Rohde, M. Bonaffini, M. Giraud, and A. Piccini, "Determinación de un índice de rendimiento académico general para medir el riesgo de deserción universitaria," *Revista de la facultad de ciencias económicas - UNNE*, no. 20, pp. 109-121, 2018.
- [95] A. Hernández, "Percepciones sobre discapacidad: Implicaciones para la atención.," *Revista Electrónica Educare (Educare Electronic Journal)*, vol. 22, no. 3, pp. 1-24, julio 2018.
- [96] H. Hamidi and A. Chavoshi, "Análisis de los factores esenciales para la adopción del aprendizaje móvil en la educación superior: Un caso de estudio de los estudiantes de la Universidad Tecnológica," *Telemática e Informática*, julio 2018.
- [97] L. Lopez and A. Beltrán , "La deserción en estudiantes de educación superior: tres percepciones en estudio, alumnos, docentes y padres de familia.," *Pista Educativas*, vol. 39, no. 126, 2017.
- [98] C. Castaño and Y. G. Gallego, "Análisis de las causas de deserción universitaria en la UNAD Udr Cali y su incidencia en los aspectos financieros para la cohorte del año 2018-2 versus 2018-1.," *Recuperado de: <https://repository.unad.edu.co/handle/10596/31691>*, 2020.
- [99] D. M. Gudelo, C. P. Casadiegos, and D. L. Sánchez, "Características de ansiedad y depresión en estudiantes universitarios," *International Journal of Psychological Research 2008.*, vol. Vol. 1, 2008.
- [100] V. Pérez, M. Valenzuela, A. Díaz, J. A. González, and J. C. Núñez, "Dificultades de aprendizaje en estudiantes universitariosde primer año.,"

SCIELO, 2013.

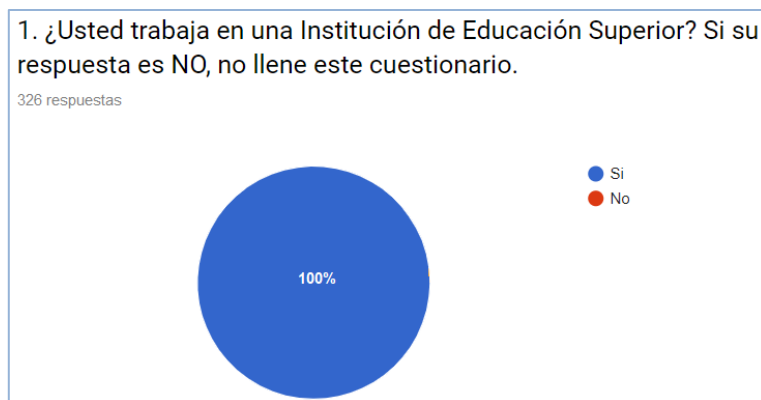
- [101] Mirza Cilimkovic, "Neural Networks and Back Propagation Algorithm," *Institute of Technology Blanchardstown*.
- [102] José Abiel Caballero Hernández, Martha Díaz Salazar, Meyli Moradillos Paz Lago, and Sonnia Pavoni Oliver, "Implementación de la Función Sigmoidal Logarítmica en un FPGA," *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 35, no. 2, mayo 2014.
- [103] B.D. Ripley, "Pattern recognition and neural networks," Cambridge, 1996.
- [104] S. Barrie, P. Ginns, and M. Prosser, "Early impact and outcomes of an institutionally aligned, student focused learning perspective on teaching quality assurance," *Assessment & Evaluation in Higher Education*, 2005.
- [105] P. McLaughlin and N Simpson, "The common first year programme: some lessons from a construction science course," *Teaching in Higher Education*. doi:10.1080/13562510601102057 , 2007.
- [106] P. Creed, T. Fallon, and M. Hood, "The relationship between career adaptability, person and situation variables, and career concerns in young adults," *Journal of Vocational Behavior*. doi:10.1016/j.jvb.2008.12.004 , 2009.

ANEXOS

Anexo 1

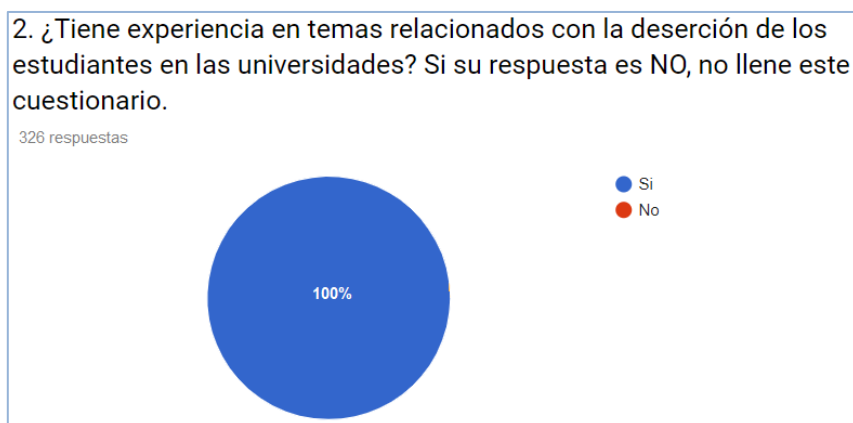
Análisis descriptivo de la primera sección del instrumento de investigación.

Sección 1: Caracterización de las instituciones y de la persona que llenan el cuestionario.



*Gráfico 64: Si trabaja en una institución Pública
Elaborado por: Investigador*

Análisis e interpretación: Al visualizar el gráfico 64, se llega a la conclusión que los 326 encuestados trabajan en la Institución Superior, es decir el 100% de la muestra.



*Gráfico 65: Experiencia con deserción estudiantil universitaria.
Elaborado por: Investigador*

Análisis e interpretación: Al interpretar el gráfico 65, se llega a la conclusión que los 326 encuestados tiene experiencia en la deserción de estudiantes universitarios, así representado el 100% de la muestra, es decir que dicha muestra fue seleccionada de manera adecuada para poder realizar el presente trabajo de investigación.

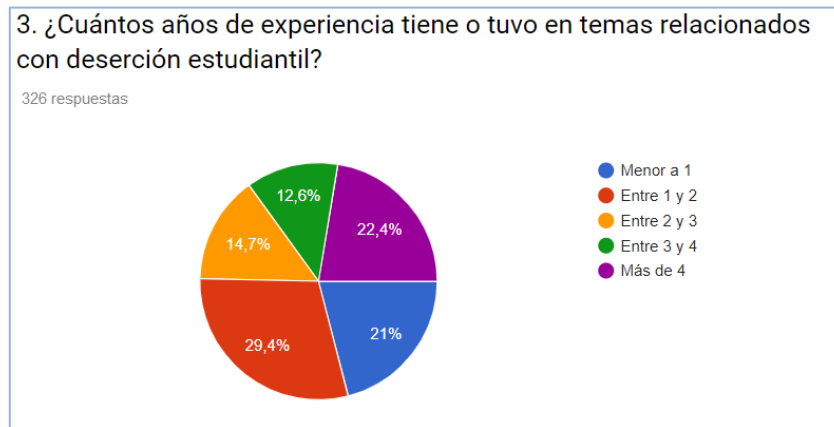


Gráfico 66: Años de experiencia con deserción estudiantil universitaria
Elaborado por: Investigador

Análisis e interpretación: Al analizar el gráfico 66, se puede concluir que en mayor porcentaje (29,4) de la muestra tiene entre 1 y 2 años de experiencia en el tema, y con un porcentaje muy cercano (22,4 %) tienen más de 4 años de experiencia, al mismo tiempo entre 3 y 4 años el 12%, siendo una muestra muy rica en conocimiento de la deserción estudiantil.

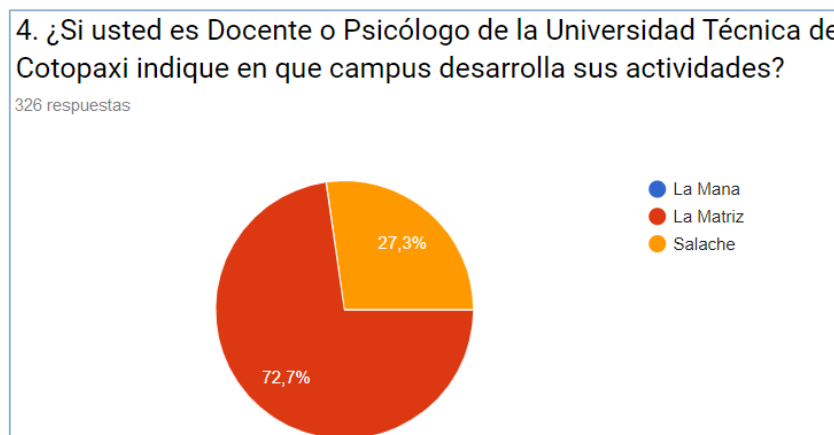


Gráfico 67: En que campus desarrolla sus actividades.
Elaborado por: Investigador

Análisis e interpretación: Al analizar el gráfico 67, que puede observar que el 72,7% de la muestra desarrolla sus actividades en el Capuz de La Matriz y un 27,3% labora en el Campus Salache.

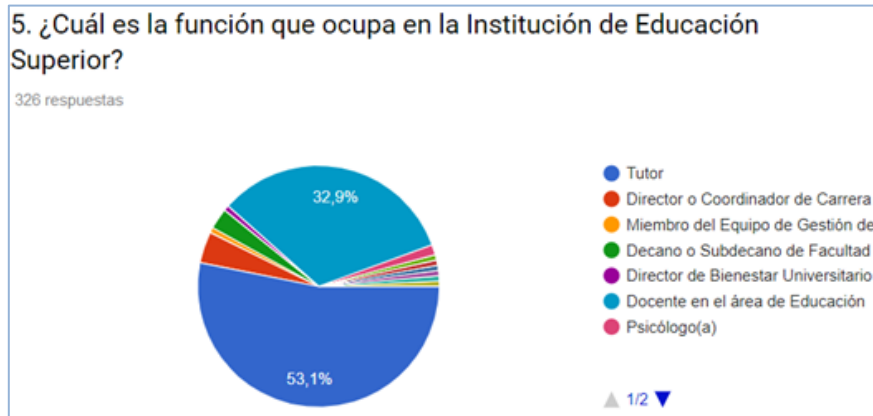


Gráfico 68: Función que ocupa en la Institución.
Elaborado por: Investigador

Análisis e interpretación: Analizando el gráfico 68, se puede determinar que en un 53.1% son Tutores y el 32,9% son docentes del área de educación, en estas dos funciones se ve reflejado el mayor porcentaje de la muestra seleccionada.

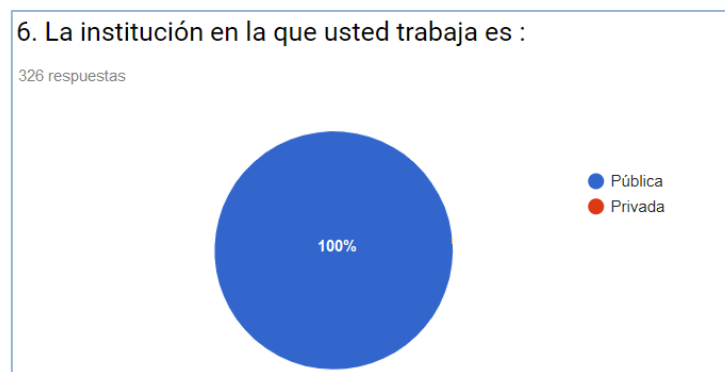


Gráfico 69: Institución Pública o Privada.
Elaborado por: Investigador

Análisis e interpretación: Interpretando el gráfico 69, se puede concluir que el 100% de la nuestra trabaja en una Institución Pública.

7. ¿Cuál es su profesión?

326 respuestas

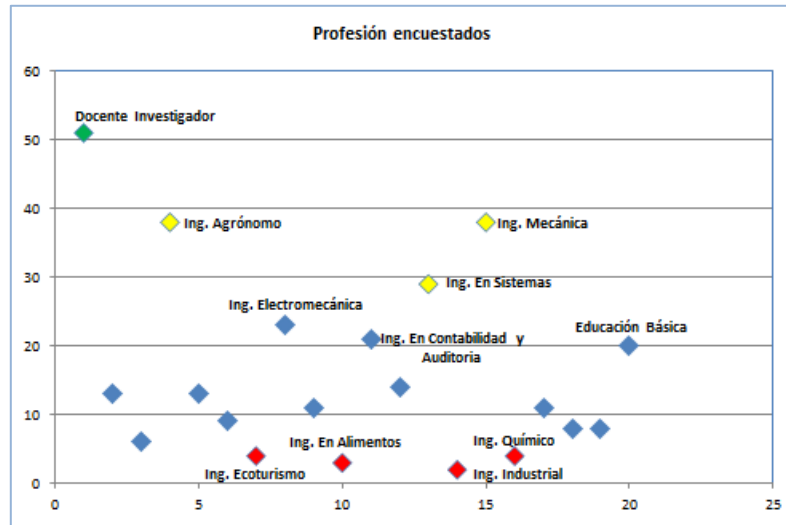


Gráfico 70: Profesión de la muestra encuestada.
Elaborado por: Investigador

Análisis e interpretación: Analizando el gráfico 70, se puede observar que en su gran mayoría de los encuestados son Docentes Investigadores, Ing. Agrónomo, Ing. Mecánico, Ing. En Sistemas y en su minoría tiene la profesión de Ing. Industrial, Ing. Ecoturismo, Ing. Químico e Ing. En Alimentos.

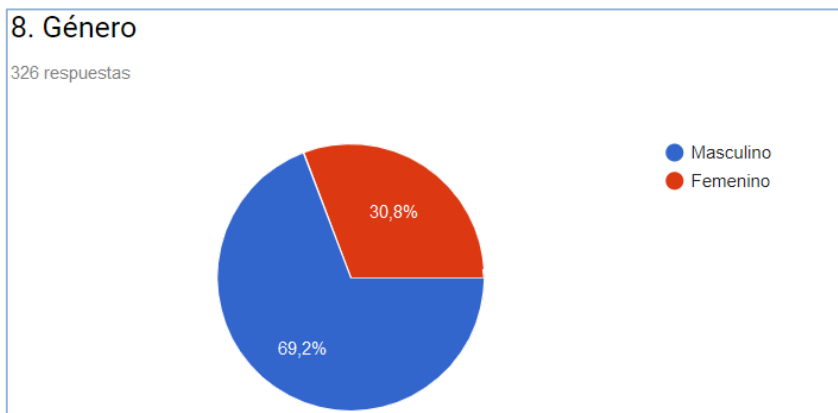


Gráfico 71: Género de los encuestados.
Elaborado por: Investigador

Análisis e interpretación: Interpretando la Imagen 71, se llega a la conclusión que en un 69.2% de los encuestados son de género Masculino y el 30.8% son de género Femenino.

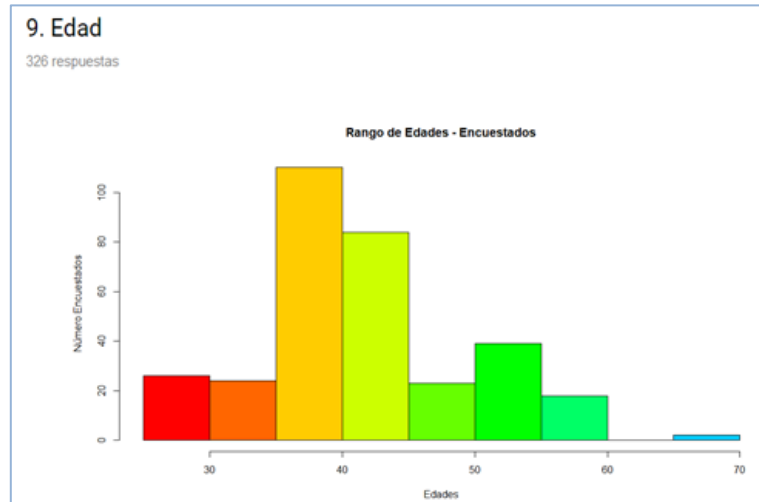


Gráfico 72: Rango de edad de los encuestados.
Elaborado por: Investigador.

Análisis e interpretación: Analizando el gráfico 72 representada por un histograma se puede observar claramente que la edad de los encuestados, en un gran porcentaje oscila entre los 35 a 45 años edad y el otro porcentaje entre 45 a 60 años.

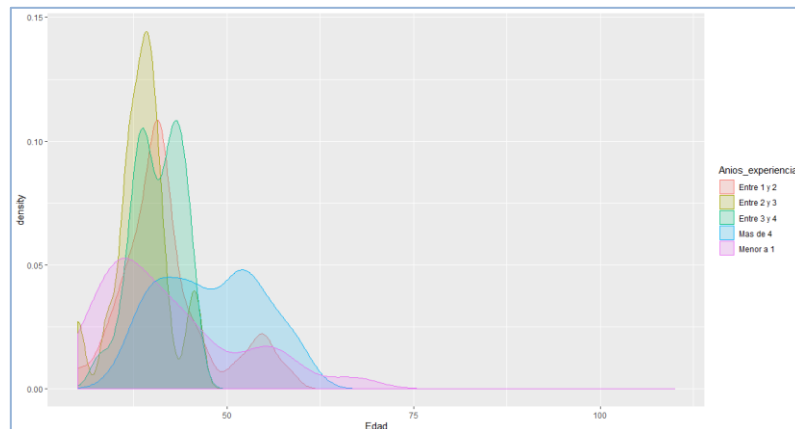


Gráfico 73: Distribución normal de la edad y años de experiencia.
Elaborado por: Investigador.

Prueba **t de Student** utilizando la librería “tidyverse” y el package “ggplot” en R, para verificar la distribución normal de dos variables en este caso son: Edad y Años de experticia en la deserción. Como podemos observar en el gráfico 73, los expertos que están entre los 35 y 45 años tienen de 1 a 4 años de experiencia en deserción universitaria.

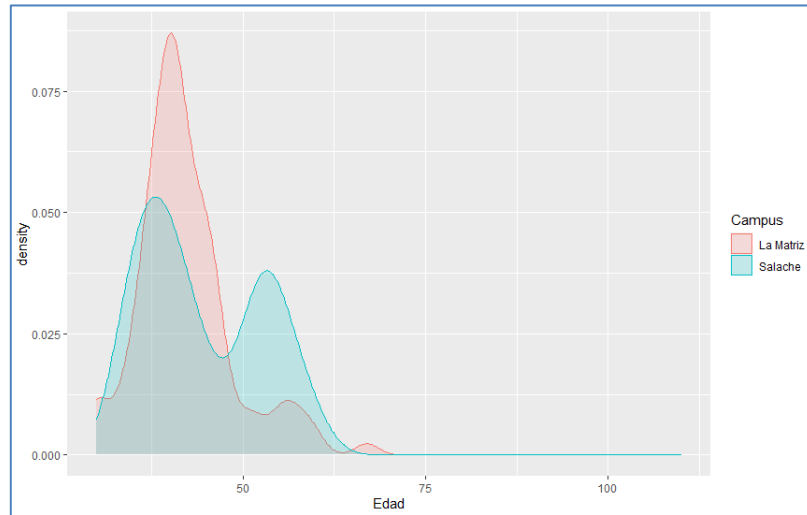


Gráfico 74: *Distribución normal de las edades de expertos en los campus de la Universidad Técnica de Cotopaxi.*

Elaborado por: Investigador.

Seguimos verificando la distribución normal de dos variables en este caso son: Edad y Campus de la Universidad donde laboran los expertos. Como podemos observar en el gráfico 74, tenemos más expertos en el Campus de la Matriz que tienen una edad de 35 a 45 años y en el Campus de Sache las edades de los expertos en su mayoría están entre los 35 a 45 años y en una menor cantidad la edad de los expertos está entre los 50 y 65 años.

Anexo 2

```

# Estructura del Conjunto de datos
str(dataset_Hijos)
entrenamiento_hijos_nrm <- entrenamiento
test_hijos_nrm <- test
View(entrenamiento_hijos_nrm)
View(test_hijos_nrm)
# FORMULA
# -----
nms <- names(entrenamiento_hijos_nrm)
ecuacion <- as.formula(paste("as.numeric(FD) ~", paste(nms[!nms %in% "FD"], collapse = " + ")))
str(nms)
str(ecuacion)
# MODELO
# -----
# especifica una primera capa oculta con 5 neuronas y una segunda capa oculta con 3 neuronas
# indica que las iteraciones se detendran cuando el "Cambio" del error sea menor a 5% entre una iteracion de op
# refiere al algoritmo "Resilient Backpropagation", que actualiza los pesos considerando unicamente el signo del
modelo.nnH <- neuralnet(ecuacion,
  data      = entrenamiento_hijos_nrm,
  hidden   = c(5,3),
  threshold = 0.01,
  algorithm = "rprop+"
)
modelo.nnH
# PREDICCION
# -----
pr.nn <- compute(modelo.nnH,within(test_hijos_nrm,rm(FD)))
View(pr.nn)
# se transforma el valor escalar al valor nominal original
des.predict <- pr.nn$net.result*(max(dataset_Hijos$FD)-min(dataset_Hijos$FD))+min(dataset_Hijos$FD)
des.real <- (test_hijos_nrm$FD)*(max(dataset_Hijos$FD)-min(dataset_Hijos$FD))+min(dataset_Hijos$FD)
View(des.predict)
View(des.real)
# SUMA DE ERROR CUADRATICO
# -----
(se.nn <- sum((des.real - des.predict)^2)/nrow(test_hijos_nrm))
#GRAFICOS
# -----
# Errores
qplot(x=des.real, y=des.predict, geom=c("point","smooth"), method="lm",
  main=paste("Real Vs Prediccion. Summa de Error Cuadratico=", round(se.nn,2)))
plot(modelo.nnH, rep="best")

```

Gráfico 75: Modelo de deserción UTC - factores padres de deserción estudiantil.
Elaborado por: Investigador.

```

# Estructura del Conjunto de datos
str(datosPadres)
entrenamiento_padres_nrm <- entrenamiento
test_padres_nrm <- test
View(entrenamiento_padres_nrm)
View(test_padres_nrm)
# FORMULA
# -----
nms <- names(entrenamiento_padres_nrm)
ecuacion <- as.formula(paste("as.numeric(FD) ~", paste(nms[!nms %in% "FD"], collapse = " + ")))
str(nms)
str(ecuacion)
# MODELO
# -----
# especifica una primera capa oculta con 5 neuronas y una segunda capa oculta con 3 neuronas
# indica que las iteraciones se detendran cuando el "Cambio" del error sea menor a 5% entre una iteracion de
# refiere al algoritmo "Resilient Backpropagation", que actualiza los pesos considerando unicamente el signo de
modelo.nn <- neuralnet(ecuacion,
  data      = entrenamiento_padres_nrm,
  hidden   = c(5,3),
  threshold = 0.04,
  algorithm = "rprop+"
)
modelo.nn
# PREDICCION
# -----
pr.nn <- compute(modelo.nn,within(test_padres_nrm,rm(FD)))
pr.nn
# se transforma el valor escalar al valor nominal original
des.predict <- pr.nn$net.result*(max(datosPadres$FD)-min(datosPadres$FD))+min(datosPadres$FD)
des.real <- (test_padres_nrm$FD)*(max(datosPadres$FD)-min(datosPadres$FD))+min(datosPadres$FD)
des.predict
des.real
# SUMA DE ERROR CUADRATICO
# -----
(se.nn <- sum((des.real - des.predict)^2)/nrow(test_padres_nrm))
#GRAFICOS
# -----
# Errores
qplot(x=des.real, y=des.predict, geom=c("point","smooth"), method="lm",
  main=paste("Real Vs Prediccion. Summa de Error Cuadratico=", round(se.nn,2)))
plot(modelo.nn, rep="best")

```

Gráfico 76: Modelo de deserción UTC - factores hijos de deserción estudiantil.
Elaborado por: Investigador.

Anexo 3

```
86 set.seed(3141592)
87 subC_datos <- createFolds(datos_padres$FD, k=10)
88
89
90 Cross_V_RedNeuronal <- function(x)
91 {
92   entrenamiento <- datos_padres_norm[-x,]
93   test <- datos_padres_norm[x,]
94   ann_model <- neuralnet(FD_0+FD_1~., data=entrenamiento, hidden=c(5,5), lifestage = "minimal", linear.output = FALSE, rep =1)
95   ann_pred <- compute(ann_model,test[,3:22])
96   ann_pred_round <- as.data.frame(round(ann_pred$net.result))
97   predic <- max.col(ann_pred_round)
98   test_res <- max.col(test[,1:2])
99   myconfusion <- confusionMatrix(table(test_res, predic))
100 # el rango [1:2] corresponde a los índices donde se encuentran las variables Accuracy y Kappa
101 #dentro de nuestra matriz de confusión
102   return (myconfusion$overall[1:2])
103 }
104
105 Cross_V_Result <- lapply(subC_datos, Cross_V_RedNeuronal)
106 Ccros_V_result_FD <-as.data.frame(Cross_V_Result)
107 Ccros_V_result_FD
108 rowMeans(Ccros_V_result_FD)
109
110 plot(ann_model, rep="best")
111
```

Gráfico 77: Código de cross_validation - factores padres de deserción estudiantil.
Elaborado por: Investigador.

```
86 set.seed(3141592)
87 subC_datos <- createFolds(datos_hijos$FD, k=10)
88
89
90 Cross_V_RedNeuronal <- function(x)
91 {
92   entrenamiento <- datos_hijos_norm[-x,]
93   test <- datos_hijos_norm[x,]
94   ann_model <- neuralnet(FD_0+FD_1~., data=entrenamiento, hidden=c(5,5), lifestage = "minimal", linear.output = FALSE, rep =1)
95   ann_pred <- compute(ann_model,test[,3:71])
96   ann_pred_round <- as.data.frame(round(ann_pred$net.result))
97   predic <- max.col(ann_pred_round)
98   test_res <- max.col(test[,1:2])
99   myconfusion <- confusionMatrix(table(test_res, predic))
100 # el rango [1:2] corresponde a los índices donde se encuentran las variables Accuracy y Kappa
101 #dentro de nuestra matriz de confusión
102   return (myconfusion$overall[1:2])
103 }
104
105 Cross_V_Result <- lapply(subC_datos, Cross_V_RedNeuronal)
106 Ccros_V_result_FD <-as.data.frame(Cross_V_Result)
107 Ccros_V_result_FD
108 rowMeans(Ccros_V_result_FD)
109
110 plot(ann_model, rep="best")
111
```

Gráfico 78: Código de cross_validation - factores hijos de deserción estudiantil.
Elaborado por: Investigador.