



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

**Contribuciones a la predicción de la deserción
universitaria a través de minería de datos**

TESIS

Para optar el Grado Académico de Doctor en Ingeniería de
Sistemas e Informática

AUTOR

Mayra Susana ALBÁN TAIPE

ASESOR

David Santos MAURICIO SÁNCHEZ

Lima, Perú

2019



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Albán, M. (2019). *Contribuciones a la predicción de la deserción universitaria a través de minería de datos*. Tesis para optar grado de Doctor en Ingeniería de Sistemas e Informática. Unidad de Posgrado, Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima, Perú.

1. Código ORCID del Autor:

<http://orcid.org/0000-0003-1519-4023>

2. Código ORCID del Asesor:

<https://orcid.org/0000-0001-9262-626X>

3. Grupo de Investigación:

Inteligencia Artificial

4. Institución que Financia la Investigación:

Universidad Técnica de Cotopaxi

5. Ubicación Geográfica: Avenida Simón Rodríguez s/n Barrio El

Ejido, Sector San Felipe.

6. Año o rango de años que abarco la investigación:

2014 - 2019

7. Cédula de Identidad: 0502311988



Universidad Nacional Mayor de San Marcos
 Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
 Unidad de Posgrado

SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE DOCTOR EN INGENIERÍA DE SISTEMAS E INFORMÁTICA

En la Ciudad Universitaria, a los veintidós (22) días del mes de mayo del 2019, siendo las 19:40 horas, se reunieron en el Auditorio de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos, el Jurado de Tesis conformado por los siguientes docentes:

- Dr. Frank Edmundo Escobedo Bailón (Presidente)
- Dr. Erik Alex Papa Quiroz (Miembro)
- Dr. Jose Carlos Daniel Alvarez Merino (Miembro)
- Dr. David Santos Mauricio Sánchez (Asesor)

Se inició la Sustentación invitando a la candidata a Doctor **Mayra Susana Albán Taipe**, para que realizara la exposición oral y pública de la tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática, siendo la Tesis intitulada:

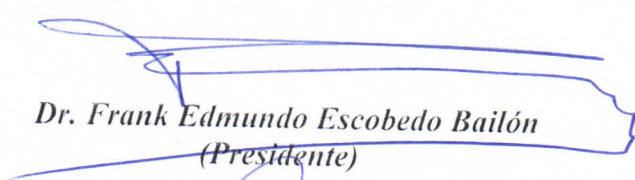
“Contribuciones a la Predicción de la Deserción Universitaria a través de Minería de Datos”

Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por la graduanda; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

veinte (20)

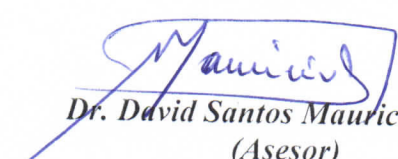
Por tanto el Presidente del Jurado, de acuerdo al Reglamento de Grados y Títulos, otorga a la Magister **Mayra Susana Albán Taipe** el Grado Académico de Doctor en Ingeniería de Sistemas e Informática.

Siendo las 20:43 horas, el Presidente del Jurado de Tesis da por concluido el acto académico de Sustentación de Tesis.


 Dr. Frank Edmundo Escobedo Bailón
 (Presidente)


 Dr. Erik Alex Papa Quiroz
 (Miembro)


 Dr. Jose Carlos Daniel Alvarez Merino
 (Miembro)


 Dr. David Santos Mauricio Sánchez
 (Asesor)



AGRADECIMIENTO

Quiero expresar mi agradecimiento al Doctor David Mauricio Sánchez, por su acertada dirección para el desarrollo de la presente investigación. A los docentes de la Universidad Nacional Mayor San Marcos por los conocimientos compartidos. A mis padres, hermanos por su apoyo incondicional y especialmente a mi hija Xiomara por su amor incondicional. También, dejo constancia de mi reconocimiento a la Universidad Técnica de Cotopaxi por contribuir en mi formación profesional.

Mayra Albán

DEDICATORIA

- A mis padres Jorge y Susana
- A mi hija Xiomara

Mayra Albán

ÍNDICE GENERAL

AGRADECIMIENTO.....	III
DEDICATORIA	IV
ÍNDICE GENERAL.....	V
ÍNDICE DE TABLAS	X
ÍNDICE DE FIGURAS.....	XIII
RESUMEN.....	XV
ABSTRACT	XVII
Capítulo I: Introduccción	1
1.1 Situación Problemática	1
1.2 Formulación del Problema.	4
1.2.1 Problema General.....	5
1.2.2 Problemas Específicos.	5
1.3 Justificación Teórica.	5
1.4 Justificación Práctica.....	6
1.5 Objetivos	9
1.5.1 Objetivo General	9
1.5 Objetivos Específicos.....	10
1.6 Hipótesis.....	11
1.6 Propuesta	11
1.6 Organización de la Tesis	11
Capítulo II: Marco Teórico	13

2.1 Deserción estudiantil.....	13
2.2 Tipos de deserción.....	14
2.3 Factores de deserción.....	15
2.3.1 Tipos de dimensiones para clasificar factores de deserción	15
2.4. Modelo de predicción de la deserción	16
2.4.1 Conceptualización de modelo.....	16
2.4.2 Conceptualización de predicción.....	17
2.4.3 Conceptualización modelo de predicción de la deserción univeritaria.....	17
Capítulo III: Estado del Arte.....	19
3.1 Introducción.....	19
3.2 Metodología de Investigación.....	21
3.2.1 Planificación de la Revisión.....	22
3.2.2 Conducción de la Revisión	23
3.3 Resultados.	24
3.3.1 Q1: ¿Cuáles son los factores usados para predecir la deserción universitaria?.....	26
3.3.2 Q2: ¿Cuáles son las técnicas utilizadas para el preprocesamiento de datos?	33
3.3.3 Q3: ¿Cuáles son las técnicas para la extracción de factores?	34
3.3.4 Q4: ¿Cuáles son las técnicas utilizadas para la predicción de la deserción universitaria?.....	36
3.4 Discusión.....	45

3.5 Conclusiones	46
Capítulo IV: Factores de deserción desde la perspectiva del estudiante.....	48
4.1 Introducción	48
4.2 Trabajos relacionados.	49
4.3 Metodología	55
4.3.1 Nuevos factores identificados.....	55
4.3.2 Recopilación de información.....	62
4.4 Resultados	63
4.4.1 Análisis descriptivo de la población	63
4.4.2 Fiabilidad de los datos.....	64
4.4.2 Modelado de datos	66
4.5 Discusión.....	71
4.5 Conclusiones	44
Capítulo V. Predicción de la deserción a través de factores tecnológicos.....	76
5.1 Introducción	76
5.2 Metodología.	78
5.2.1 Conjunto de Datos.....	79
5.2.2. Factores Tecnológicos.....	80
5.3 Resultados.	80
5.4 Discusión y Conclusiones	82
Capítulo VI: Árbol de decisión para la identificación temprana de estudiantes universitarios en riesgo de deserción	84

6.1 Introducción.	84
6.2 Trabajos relacionados.	85
6.3 Materiales y métodos.	86
6.3.1 Conjunto de datos.....	86
6.3.2 Análisis de datos	87
6.4 Resultados y discusión	87
6.5 Conclusiones	91
Capítulo VII: Red Neuronal para predecir la deserción en las universidades	93
7.1 Introducción	93
7.2 Trabajos relacionados	94
7.3 Materiales y métodos	95
7.3.1 Conjunto de datos.....	95
7.3.2 Análisis de datos	96
7.4 Resultados y discusión	97
7.4.1 Preprocesamiento de datos.....	97
7.4.2 Predicción de la deserción en las universidades	98
7.5 Conclusiones	104
Capítulo VIII. Meta clasificador para la predicción de la deserción estudiantil universitaria	106
8.1 Introducción	106
8.2 Materiales y métodos	108
8.2.1 Conjunto de datos	108
8.2.2 Análisis de datos	108

8.2.3 Análisis de los métodos	109
8.3 Resultados y discusión	110
8.3.1 Preprocesamiento de los datos	110
8.3.2 Extracción de factores	110
8.3.3 Predicción de la deserción	111
Capítulo IX: Conclusiones, limitaciones y trabajos futuros	116
9.1 Conclusiones	116
9.1 Limitaciones	117
9.1 Trabajos futuros	118
9.1 Publicaciones	118
Referencias Bibliográficas	120
Anexos.....	136
Anexo A. Encuesta.....	136
Anexo B. Producción científica.....	140

ÍNDICE DE TABLAS

Tabla 1. Gasto por educación.....	7
Tabla 2. Efectos de la deserción estudiantil.....	9
Tabla 3. Criterios para la selección de documentos.....	23
Tabla 4. Producción científica seleccionada.....	24
Tabla 5. Factores de la dimensión personal	26
Tabla 6. Factores de la dimensión académica.....	28
Tabla 7. Factores de la dimensión económica	30
Tabla 8. Factores de la dimensión social	31
Tabla 9. Factores de la dimensión institucional.....	33
Tabla 10. Técnicas para el preprocesamiento de los datos	34
Tabla 11. Técnicas para la extracción de factores	35
Tabla 12. Técnicas de inteligencia artificial	36
Tabla 13. Técnicas estadísticas	38
Tabla 14. Presición de las técnicas estadísticas	40
Tabla 15. Presición de las técnicas de inteligencia artificial	40
Tabla 16. Herramientas utilizadas por las técnicas de inteligencia artificial	43
Tabla 17. Herramientas utilizadas por las técnicas estadísticas	44
Tabla 18. Dimensión institucional	50
Tabla 19. Dimensión personal	50
Tabla 20. Dimensión económica.....	52
Tabla 21. Dimensión académica	52

Tabla 22. Dimensión social.....	54
Tabla 23. Nuevos factores de deserción universitaria identificados	61
Tabla 24. Estadística descriptiva.....	64
Tabla 25. Resultado de la validez de la encuesta	65
Tabla 26. Resultado de la valoración de la escla de Likert	65
Tabla 27. Medidas de tendencia central sobre la percepción de la deserción	66
Tabla 28. Coeficiente de significancia de los factores	67
Tabla 29. Test de Hosmer - Lemershow	69
Tabla 30. Wald Test	70
Tabla 31. Probabilidad de deserción estudiantil universitaria	70
Tabla 32. Factores tecnológicos que influyen en la deserción universitaria.....	80
Tabla 33. Selección de factores.....	81
Tabla 34. Trabajos relacionados con la aplicación de árbol de decisión en la predicción de la deserción.....	86
Tabla 35. Descripción de variables	88
Tabla 36. Importancia de las variables independientes	88
Tabla 37. Resumen del modelo	89
Tabla 38. Resultados de la precisión del árbol de decisión CHAID	91
Tabla 39. Red neuronal para predecir la deserción en las universidades	94

Tabla 40. Definición de las variables	97
Tabla 41. Importancia de las variables	97
Tabla 42. Resumen del modelo mediante Perceptrón Multicapa	100
Tabla 43. Resumen del modelo mediante Función de Base Radial	101
Tabla 44. Presición de la predicción de los modelos propuestos	104
Tabla 45. Estadística descriptiva factores de estudio	109
Tabla 46. Resultados de la selección de atributos mediante el método Greedy Stepwise	111
Tabla 47. Porcentaje de presición de la predicción.....	112

ÍNDICE DE FIGURAS

Figura 1. Esquema de la propuesta	11
Figura 2. Proceso de minería de datos para predecir la deserción estudiantil universitaria	20
Figura 3. Proceso para la revisión sistemática de literatura.....	23
Figura 4. Tendencia temporal de publicaciones seleccionadas sobre la deserción estudiantil universitaria	24
Figura 5. Publicaciones sobre la predicción de la deserción universitaria por tipo de fuente	25
Figura 6. Modelo conceptual para determinar factores de deserción	62
Figura 7. Residuos OLS del modelo de regresión logística.....	68
Figura 8. Probabilidad de deserción $x=1$	72
Figura 9. Probabilidad de deserción $x=3$	73
Figura 10. Probabilidad de deserción cuando todos los factores se activan a la vez	74
Figura 11. Proceso de predicción de la deserción universitaria..	79
Figura 12. Modelos de regresión logística y árbol de decisión ..	82
Figura 13. Modelo de árbol de decisión.....	90
Figura 14. Modelo de red neuronal.....	96
Figura 15. Curva de elevación variable dependiente (Des) modelo Perceptrón Multicapa	100
Figura 16. Curva de elevación variable dependiente (Des) modelo Función Base Radial	102

Figura 17. Pseudo-probabilidad predicha	103
Figura 18. Precisión de la predicción métodos Multilayer Perceptrón, Random Forest, PART, J48	113
Figura 19. Metaclasificador para la predicción de la deserción estudiantil	114

RESUMEN

La deserción es considerada como una problemática en el contexto de la educación superior. Como consecuencia negativa destaca el costo económico y social para los alumnos, las universidades y los gobiernos. La identificación temprana de los estudiantes en riesgo de desertar se ha convertido en un reto para los administradores de las universidades. La predicción de la deserción universitaria es fundamental y ha cobrado gran importancia debido a sus altas tasas que en la actualidad aún se mantienen. Su importancia radica en que esta podría convertirse en un mecanismo eficaz para la determinación de nuevas estrategias que permitan la reducción de sus tasas y una mejor calidad de vida académica para el estudiante. El impacto positivo de predecir la deserción constituye el aumento de la retención de los estudiantes, mayor presupuesto, reconocimiento institucional y consecución de sus objetivos, tanto académicos como de gestión para las universidades. La revisión de la literatura permitió identificar diversos modelos de predicción para determinar los alumnos en riesgo de desertar. Sin embargo, los estudios analizados en general no consideran una revisión integral de literatura sobre la predicción de la deserción universitaria. También se pudo identificar una limitada producción científica que analiza factores de deserción desde la perspectiva del estudiante, que es el actor principal de la deserción, y la construcción de modelos híbridos de predicción que permitan comprender mejor manera el problema de la deserción en las universidades. Por consiguiente, el objetivo de la investigación consiste en contribuir al proceso de predicción de la deserción estudiantil universitaria a través del estudio integral de factores, técnicas y herramientas de minería de datos usados con este fin. Se concluye que la predicción de la deserción en las universidades puede variar, ya que dependerá de los factores de ingreso, del contexto educativo estudiado, del entorno de educación aplicado, y de los antecedentes de los estudios para los que fueron usados. Por otro lado, se considera importante determinar si es suficiente con predecir la deserción o si se requiere incorporar estudios que

establezcan estrategias para mitigar la deserción en las instituciones de educación superior.

Keywords: deserción estudiantil universitaria, factores de deserción, predicción de la deserción, minería de datos, machine learning.

ABSTRACT

Student desertion is a problem that affects universities, with negative consequences for the social and economic cost for students, higher education institutions and governments. Early identification of students at risk of dropping out has become a challenge for universities. The prediction of the university dropout is fundamental and has become important in the context of higher education because of the high dropout rates that still exist today. The importance of the prediction of the university student desertion is that it could become an effective mechanism for the determination of new strategies that allow the reduction of dropout rates and a better quality of academic life for the student. The positive impact of predicting dropouts is the increase in student retention, greater budget, institutional recognition and achievement of their academic and management objectives for universities. The review of the literature made it possible to identify various methods and models developed to help the universities in the prediction of students at risk of dropping out. However, the studies analyzed in general do not consider a comprehensive review of literature on the prediction of attrition in universities. The determination of factors from the perspective of the student who is the main actor of the desertion and the construction of hybrid models of prediction that allow to better understand the problem of the desertion in the universities. Therefore, the objective of the research is to contribute to the prediction process of university student desertion through the integral study of factors, techniques and tools of data mining used to learn the desertion in the universities. It is concluded that the prediction of the desertion in the universities can vary depending on the factors of income, the educational context studied, the environment of applied education, and the background of the studies for which they were used. On the other hand, it is important to determine if it is enough to predict the dropout or if it is necessary to incorporate studies that establish strategies to mitigate the dropout in higher education institutions.

Keywords: university student desertion, factors of desertion, prediction of the desertion, mining of data, machine learning.

CAPÍTULO I: INTRODUCCIÓN

1.1 Situación Problemática

La problemática de la deserción estudiantil afecta negativamente a las Instituciones de Educación Superior (Elias, 2008), tema que ha sido estudiado desde hace varias décadas y utilizado como argumento de gran importancia dentro de las políticas públicas de la educación universitaria en varios países alrededor del mundo. El problema de la deserción en las universidades ha sido estudiado a partir de 1920 (Sangodiah, 2015), en donde varios investigadores han expresado su preocupación sobre las altas tasas de deserción presentes en las universidades.

Varias investigaciones se han propuesto en la literatura para definir la deserción, establecer sus causas y predecir los estudiantes en riesgo de desertar. Tinto (1975) en su estudio, propone un modelo teórico que describe a la deserción estudiantil como un proceso longitudinal de interacciones en las que se relacionan la integración académica y social del alumno con objetivos del compromiso personal e institucional que conllevan a determinar las decisiones de abandono por parte de los alumnos. Según el autor, la deserción puede estar relacionada con la falta de congruencia entre el clima intelectual del estudiante y su ambiente social que se genera en la institución.

Para Kaufman (2004), la deserción tiene consecuencias sociales en términos de las expectativas de los estudiantes y sus familias, así como también consecuencias emocionales por la disonancia entre las aspiraciones de los jóvenes y sus logros. Los efectos negativos causados por esta problemática afectan al alumno en su entorno social, económico y emocional.

Por otro lado, las altas tasas de deserción generan también efectos negativos para las universidades, su impacto puede verse reflejado en la disminución de su presupuesto, la disminución de las tasas de graduación y el alto costo en los procesos administrativos y académicos. (Krstic, 2016)

Para Tan (2015), la premisa para reducir las tasas de abandono universitario es conocer sus causas o factores que estén asociados a la deserción, sin embargo, esto se ha convertido en una tarea difícil para los investigadores debido a que la deserción es considerada como multifactorial y se basa en las diferencias individuales de los estudiantes en función del sistema de enseñanza universitario aplicado que, en muchas ocasiones, difieren del contexto social en el que el estudiante se desenvuelve.

Cabe destacar que algunos de los estudios propuestos en la revisión de la literatura, relacionados con el tema objeto de estudio, han considerado a la deserción como una problemática educativa con un elevado impacto económico y social tanto para las instituciones de educación superior como para los gobiernos. Por lo tanto, se considera que una de las razones por las cuales este problema aún no ha sido resuelto es porque no se ha determinado con exactitud las causas que lo generan.

Tinto (1975), señala que el problema de la deserción en las universidades está relacionado con factores propios del estudiante y de su nivel de integración con su entorno académico para alcanzar con éxito su título universitario. De manera similar, Brookover (1975), en su estudio, determina variables sociales y socio psicológicas en un esfuerzo por identificar factores que permitan explicar las causas de la deserción. Tumen (2008) señala que los factores asociados al rendimiento académico deben ser controlados para detener el abandono estudiantil en las universidades. Para Chen (2012), la deserción obedece a características institucionales, las mismas que, según el autor, contribuyen a que se generen condiciones que reduzcan los riesgos de abandono estudiantil.

En el mismo sentido, Hoffait (2017) indica que existen varias razones que pueden explicar el fracaso académico de los jóvenes en las universidades, debido a una escasa orientación que los alumnos reciben para seleccionar su campo de formación académica acorde a sus fortalezas y habilidades cognitivas. Con base en lo expuesto, se puede identificar que existe un limitado consenso entre los investigadores y el establecimiento de las causas de la deserción. Se considera importante que este tema sea tratado a profundidad debido a que un inadecuado uso de factores de deserción podría ocasionar graves dificultades en la predicción de la misma.

Por otro lado, la predicción de la deserción universitaria se torna importante a partir de 1975. Astin (1975) propone un modelo de congruencia basado en las percepciones del individuo en sí mismo y sus relaciones con otros individuos. La predicción de la deserción en las Instituciones de Educación Superior ha sido cuestionada, por sus altas tasas de deserción estudiantil que aún existen en las universidades. Sin embargo, varios trabajos han sido propuestos en la revisión de la literatura para predecir la deserción y determinar sus causas. Wray, Brett, Aspland & Gardiner (2012), Thammasiri, Delen, Meesad & Kasap (2014), Saranya & Rajeswari (2016), Aulck, Velagapudi, Blumenstock & West (2016) utilizaron técnicas de regresión logística para predecir la deserción en las universidades.

De manera similar, Márquez, *et.al.*, (2013), Herzog (2006), Kotsiantis, *et.al.*, (2003), Simpson, 2004, Lykourantzou, *et. al.*, (2009) establecen modelos de predicción de la deserción mediante la utilización de técnicas de minería de datos como las redes neuronales, árbol de decisión, técnicas de aprendizaje automático, máquinas de soporte vectorial, algoritmos de aprendizaje basados en instancias.

A pesar de las considerables investigaciones en esta área, hay una falta de claridad en torno a la estandarización de métodos que permitan la selección adecuada de factores predominantes de deserción, los cuales son utilizados como variables de ingreso a los modelos predictivos y la determinación de las

técnicas adecuadas de minería de datos que permitan obtener con adecuada precisión los alumnos que se encuentran en riesgo de desertar de las universidades.

Cabe destacar que, en su gran mayoría, los modelos aplicados para tratar este problema presentan altas tasas de error que superan el 20 % (Duarte, 2014), esto puede deberse a que, al trabajar con datos de corte transversal o longitudinal con base en la dimensionalidad de conjunto de datos, se corre el riesgo de no tratar eficientemente las variables que funcionarán como entradas a los modelos predictivos. De igual manera, el desequilibrio del conjunto de datos puede afectar a la clasificación de clases o etiquetas en la construcción precisa de un modelo de predicción.

A pesar de que se han realizado cambios sistémicos eficaces para combatir este problema, los estudiantes continúan enfrentando una crisis educativa con la mayor propensión de abandonar sus estudios universitarios (Liang, 2016). Por tal razón, es importante diseñar modelos predictivos de deserción que permitan la integración de factores y técnicas de minería de datos adecuadas para identificar con aceptable grado de precisión los estudiantes en riesgo de abandonar las aulas universitarias.

1.2 Formulación del Problema

1.2.1 Problema general

¿De qué manera el uso de la minería de datos contribuye a la predicción de la deserción estudiantil universitaria?

1.2.2 Problemas específicos

- ¿Existen nuevos factores que influyen en la predicción de la deserción estudiantil universitaria?
- ¿Cuál es el nivel de influencia de los nuevos factores identificados en la predicción de la deserción estudiantil universitaria?
- ¿Cuál es el nivel de influencia de los factores tecnológicos en la predicción de la deserción estudiantil universitaria?
- ¿Cuáles son las técnicas de minería de datos que contribuye a la predicción de la deserción estudiantil universitaria?
- ¿De qué manera el diseño de un modelo híbrido a través de minería de datos contribuye a la predicción de la deserción estudiantil universitaria?

1.3 Justificación Teórica

Para Zacharis (2014), el problema de la predicción de la deserción estudiantil universitaria ha cobrado importancia. La literatura ha permitido identificar que en las últimas décadas un incremento de producción científica está orientada a estudiar la predicción de la deserción. Sin embargo, los modelos analizados, en su mayoría, estudian la deserción cuando esta ya se ha presentado, es decir, cuando el estudiante ya abandonó sus estudios universitarios, por lo que se considera indispensable el diseño de una herramienta de predicción que permita de forma temprana la identificación de estudiantes en riesgo de abandonar las aulas universitarias. (Lykourantzou, *et.al.*, 2009)

Por otro lado, los modelos de predicción identificados que aplicaron técnicas estadísticas para sus procesos experimentales no producen datos predictivos convincentes debido a que están sujetos a pruebas sustantivas para establecer su sensibilidad, fiabilidad y validez de los instrumentos aplicados. En general,

se pudo identificar que existe limitada estandarización de instrumentos para evaluar los constructos.

Por lo expuesto, se considera necesario abordar la predicción de la deserción en las universidades desde una perspectiva conceptual y técnica amplia, de modo que se considere las diferencias dimensionales que aportan los modelos diseñados para tratar este problema y reducir su impacto.

1.4 Justificación Práctica

La finalización en la universidad no siempre ha sido la norma para la sociedad. En la década de 1940, menos de la mitad de la población estadounidense comprendida entre edades de 25 a 29 años habrían terminado la universidad. Si bien ha habido un esfuerzo concertado por cerrar la brecha de la deserción universitaria y disminuir sus tasas, investigaciones como la de Kaufman, *et.al.*, (2002) señalan que la deserción ha causado efectos en el orden económico para las universidades y los gobiernos.

La asignación adecuada de recursos financieros dedicados a la educación es una de las estrategias claves que han optado los gobiernos de algunos países para beneficiar a los estudiantes y a las universidades, Organización para la Cooperación y el Desarrollo Económico (OECD, 2008). Si se considera que el gasto en la educación es una inversión que puede ayudar a fomentar el crecimiento económico, mejorar la productividad, contribuir al desarrollo personal y social para disminuir la inequidad social (OECD, 2008), por tal razón, es importante evaluar las causas que determinan la deserción en las aulas universitarias y predecir el riesgo de abandono universitario.

A medida que aumenta el nivel de educación, se incrementa el gasto público anual por alumno. Los países que tienen un porcentaje superior de gasto medio anual por estudiante al año 2012 son Estados Unidos con un 72 %, que corresponde a 25.576 dólares, mientras que Japón presenta una cifra de 20.750 dólares, Noruega 18.512 dólares y México 7.872 dólares (Panorama de la Educación España, 2013). Al relacionar el número de estudiantes que

desertan cada año con el promedio de gasto efectuado por las universidades, se puede establecer cifras altas del gasto público por este concepto.

En América Latina, para el sistema nacional, el efecto de la deserción en términos de costos que superan los 96.2 millones de dólares por año. De acuerdo a esta cifra, el gasto por abandono es equivalente al 26 % del gasto público en la educación universitaria (Organización para las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2013). En la tabla 1, se presenta datos correspondientes al gasto por educación en algunos países. Es así que la deserción se convierte en un fenómeno que refleja un déficit de calidad en el sistema educativo y lastra tanto la rentabilidad como los retornos públicos. (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2014)

Tabla 1
Gasto en la Educación Superior

Gasto en Educación	Lugar	Gasto
La deserción estudiantil es un fenómeno que refleja un déficit de calidad en el sistema educativo.	México	7.872 USD
	Estados Unidos	25.576 USD
	Países de la OECD	13 % del gasto público
	Países de América Latina	4-5 % del producto interno bruto.
	Chile	1.9 % del producto interno bruto
	América Latina	96.2 millones por año

Fuente: Organización para el Desarrollo Económico; Organización de las Naciones Unidas, para la Educación, la Ciencia y la Cultura, (2014).

Otros de los efectos causados por la deserción son las bajas tasas de graduación. En Estados Unidos, Hungría, Noruega, Nueva Zelanda y Suecia más del 40 % de los estudiantes que inician un programa universitario no completan con éxito sus estudios en los tiempos oficiales. (Panorama de la Educación, 2013)

Según el Informe de la Educación para la Calidad, al año 2010 la tasa de deserción en Colombia fue del 45.4 %, mientras que la tasa deserción en Brasil corresponde al 53.7 % (González, 2005). En Costa Rica, la tasa de deserción corresponde al 54 % y en el Reino Unido se presentó en un 18 %. (Organización para las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2013)

Las cifras muestran que en América Latina aún queda mucho por avanzar, para el año 2010 solo uno de cada 10 jóvenes comprendidos entre las edades de 25 a 29 años de edad habría completado sus estudios universitarios. Según los patrones de graduación actuales se calcula que en los países pertenecientes a la Organización para la Cooperación y el Desarrollo Económicos (OCDE) una media de los 40 % jóvenes adultos completarán su educación terciaria. La proporción oscila entre menos de 25 % en Chile, México y Turquía, y un 50 % más en Australia, Dinamarca, Islandia, Nueva Zelanda, Polonia y Reino Unido (Panorama de la Educación, 2013). Esto conlleva a que los gobiernos de turno enfrenten una nueva crisis social por el incremento de desempleo.

La conclusión alcanzada es que parte del coste público de la educación debería ser recuperado para el sistema público, pero la magnitud del retorno esperado se ve afectado por los altos índices de fracaso universitario que pueden ser consideradas como inversiones fallidas realizadas por los gobiernos en la educación, especialmente en la educación superior, debido a las grandes cantidades del presupuesto universitario encaminadas a la infraestructura y mejora constante de la calidad de la educación. Este fenómeno reduce en muchas ocasiones los presupuestos a las universidades, debido a que estos son calculados con base en el número de estudiantes que permanecen en el sistema de educación superior.

Es así que se considera importante señalar que el beneficio de resolver este problema social es alcanzar un nivel de educación superior que beneficie a los alumnos, a las universidades y a la sociedad, además, esto conllevaría a mejorar la inequidad social a la que se enfrentan los jóvenes por no tener un

título universitario, así como también a la inserción en el campo laboral con un salario digno acorde a las competencias universitarias adquiridas.

Por otro lado, la sociedad a su vez se beneficiará a través de la reducción del gasto público en programas de bienestar social, que son aplicados para disminuir el impacto de la deserción por sus múltiples efectos negativos en la sociedad y en la economía de los países, especialmente en aquellos llamados como países en desarrollo.

Tabla 2

Efectos de la deserción estudiantil en la educación

LUGAR	EFEECTO
OECD, 2017	40 % de adultos jóvenes completará sus estudios de educación.
Chile, México y Turquía	25 % de adultos jóvenes completará sus estudios de educación.
México, Turquía	50 % de adultos jóvenes completará sus estudios de educación.
En Estados Unidos, Hungría, Noruega, Nueva Zelanda y Suecia	Más del 40 % no se gradúan en el nivel de educación terciaria.
Colombia	La tasa de deserción por cohorte es del 45.4 %.
Brasil	La tasa de deserción por cohorte es del 59 %.
Costa Rica y Chile	La tasa de deserción por cohorte es del 59 %.
Reino Unido	La tasa de deserción por cohorte es del 18 %.
Chile	La tasa de deserción global es de un 53.7 %, mientras que entre el 5 % y el 40 % de los estudiantes abandonan la Universidad.
América Latina	1 de cada 10 jóvenes de 25 a 29 años de edad había completado cinco años de educación superior en 2010.

Fuente: Organización para el Desarrollo Económico; Organización de las Naciones Unidas, para la Educación, la Ciencia y la Cultura (2017).

1.5 Objetivos

1.5.1 Objetivo General

Contribuir a la predicción de la deserción estudiantil universitaria a través del uso de técnicas de minería de datos.

1.5.2 Objetivos Específicos

- a) Identificar nuevos factores que influyen en la predicción de la deserción estudiantil universitaria, a través del estudio de las teorías administrativas y de educación.
- b) Establecer el nivel de influencia de los nuevos factores en la predicción de la deserción estudiantil universitaria.
- c) Establecer el nivel de influencia de los factores tecnológicos en la predicción de la deserción estudiantil universitaria.
- d) Diseñar modelos de minería de datos para predecir con acertado grado de precisión la deserción estudiantil universitaria.
- e) Construir un modelo híbrido de predicción de la deserción estudiantil universitaria para mejorar el nivel de precisión de las técnicas de minería de datos usadas de manera individual y mejorar su tasa de predicción.

1.6 Hipótesis

Si se determinan factores adecuados de deserción y se diseñan modelos de alto rendimiento a través del uso de minería de datos, entonces se podrá contribuir a la predicción de la deserción estudiantil universitaria, para establecer políticas y estrategias que permitan mitigar las tasas de deserción en las universidades.

1.7 Propuesta

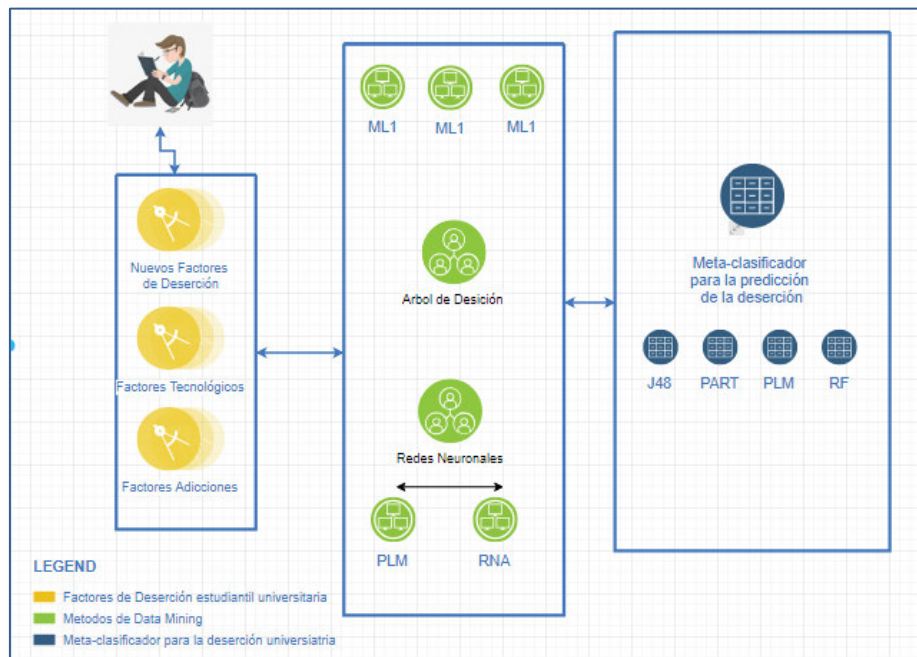


Figura 1. Esquema de la propuesta de investigación

Fuente: Elaborado por el autor

En esta sección se presenta la propuesta de investigación, la misma que se clasifica en las siguientes etapas:

- 1.- La revisión integral sistemática de literatura sobre predicción de la deserción estudiantil universitaria
- 2.- La determinación de nuevos factores de deserción en las universidades
- 3.- Diseño de modelos de predicción a través de técnicas de minería de datos
- 4.- Diseño del modelo híbrido de predicción de la deserción estudiantil universitaria

1.8 Organización de la Tesis

La presente investigación está estructurada en nueve capítulos, los cuales se describen a continuación.

En los capítulos dos y tres, se presenta el marco teórico de la investigación y la revisión integral de literatura de la predicción de la deserción estudiantil universitaria respectivamente, la cual involucra el análisis de factores, técnicas y herramientas de minería de datos usados para tratar de resolver el problema de la deserción.

En el capítulo cuatro, con base en la revisión de la literatura, se muestra el estudio de las teorías administrativas y de educación que dieron como resultado la identificación de 11 nuevos factores de deserción desde la perspectiva del estudiante y los expertos, las mismas que fueron utilizadas como variables predictoras en el modelo de predicción propuesto.

En el capítulo cinco, se presenta dos modelos de predicción basados en regresión logística y árbol de decisión sobre la base de factores tecnológicos que influyen negativamente en la deserción estudiantil universitaria.

En el capítulo seis, se presenta un árbol de decisión para identificación temprana de estudiantes en riesgo de desertar de las universidades sobre la base de factores relacionados con las adicciones de los alumnos.

En el capítulo siete, se presenta el diseño de red neuronal basada en dos métodos: perceptrón multicapa y función de base radial sobre la base de los nuevos factores de deserción identificados en el capítulo 2.

En el capítulo ocho, se desarrolla el modelo híbrido de predicción mediante el uso de métodos de minería de datos.

En el capítulo nueve, se presentan las conclusiones, limitaciones y trabajos futuros que podrían ser desarrollados sobre la base de la presente investigación.

CAPÍTULO II: MARCO TEÓRICO

En este capítulo se definen los conceptos que sirven de complemento de la investigación tales como: la conceptualización de la deserción, tipos de deserción, conceptualización respecto a modelos de predicción deserción.

2.1 Deserción estudiantil

La literatura permitió determinar varias definiciones sobre la deserción estudiantil universitaria.

La deserción es considerada en el ámbito de la educación superior como el abandono de los estudios establecida por diferentes causas y como un fenómeno de retiro forzoso de los estudiantes en el sistema de educación superior (Díaz, 2008).

Para Páramo (1999) la deserción está relacionada con el abandono definitivo del contexto de la educación universitaria respecto a su formación académica y que en muchas ocasiones ha sido relacionada con el término de mortalidad estudiantil.

Himmel (2002) abandono prematuro de un programa antes de alcanzar un título o grado. Se considera a un estudiante desertor cuando abandona los estudios durante un periodo de tiempo suficientemente extenso como para descartar la posibilidad de reincorporarse a sus estudios.

Según Fiehegen (2005) la deserción en las universidades puede ser considerada como un proceso de retiro individual voluntario o forzoso

generado por la influencia negativa interna y externa de las características del alumno en el contexto educativo.

Rodríguez (2005) acto concreto del estudiante de interrupción de estudios o situación en la cual un estudiante inicia sus estudios, pero no los termina, siendo el abandono transitorio o definitivo.

En el mismo sentido, la deserción es considerada como un acto de abandono de la universidad y como el fracaso para completar un programa de estudios Vries (2011).

2.2 Tipos de deserción

Los tipos de deserción que se presentan de basan en el estudio de Páramo (2012) y Gutierrez, Ulloa, Simancas y Velazco (2017):

- Deserción definitiva: cuando un estudiante no retoma su formación académica.
- Deserción por factores: depende de la causal que ocasiona la separación del estudiante del sistema de educación superior.
- Deserción por cambio de facultad
- Deserción por cambio de programa académico

Cuando un alumno no se matricula por más de dos periodos académicos consecutivos, el alumno puede ser considerado como desertor (Andreu, 2008).

2.3 Factores de deserción

En referencia a las causas que influyen en la deserción estudiantil en las universidades se presentan algunas conceptualizaciones:

Canales (2007) los factores son considerados como las causas que originan el problema de la deserción en la educación superior.

Vélez (2006) características del estudiante y su entorno que pueden influir en la decisión de abandonar las aulas universitarias.

Para Sanchez (2002) los factores de deserción son considerados como causas diversas y complejas o razones de la deserción.

Castaño (2008) se refieren aquellos elementos, variables o características internas y externas del estudiante y su entorno académico.

2.3.1 Tipo de dimensiones para clasificar factores de deserción

En base a la literatura se identifican dimensiones para clasificar factores relacionados con la deserción de los estudiantes universitarios.

Dimensión personal: características del comportamiento del alumno que influyen en su proceso de formación académica y en su desarrollo en el entorno educativo (Yurzetulk, 2014).

Dimensión Económica: referencia a las condiciones económicas del estudiante y de su entorno familiar para comprometer recursos financieros en la educación del estudiante (Arulampalam, 2007).

Dimensión Social: Enfoca la importancia de la interacción del alumno con su entorno social; dicha interacción es el marco en el que se constituye el sujeto en su relación con la norma académica, la institución y los hábitos de estudio (Delen, 2015).

Dimensión Académica: relacionados con el desarrollo académico del alumno en su proceso formativo.

Dimensión Institucionales: Consideran las características estructurales y funcionales de una institución relacionadas con la formación académica del alumno.

2.4 Modelo de Predicción de la Deserción

2.4.1 Conceptualización de modelo

Se define modelo como el producto final de un proceso de modelización. En el modelo residen las principales características del objeto al que representa y por lo tanto se puede estudiar sobre él, los comportamientos frente a ciertos requerimientos y circunstancias (Diccionario de Informática, 2004).

Son representaciones idealizadas de la estructura real y su medio ambiente, está fuertemente influenciada por los objetivos de la computación. Los modelos representan a la realidad sólo si se toman en cuenta todos los factores que afectan a las conclusiones de ellos.

Aproximación teórica a lo real, por medio de lo cual los postulados y suposiciones conceptuales que pueden ser aplicadas a la realidad. Son estructuras simplificadas o conocidas que se emplean para investigar la naturaleza de los fenómenos que los científicos desean investigar (Navia y Tamayo, 2018).

Un modelo es una representación abstracta de un proceso en el mundo real, una descripción global de un conjunto de datos, el cual puede ser predictivo o descriptivo (Pérez, 2018).

Por lo general, los modelos describen y explican fenómenos, que se ocultan en el conjunto de datos y se pueden utilizar para predecir el valor del atributo de destino sabiendo los valores de la entrada de atributos (Azis, 2018).

2.4.2 Conceptualización de predicción

Enunciado esperado de valores en torno a una variable aún no observada, a partir del conocimiento de los valores de otra variable (Navia y Tamayo, 2018).

La predicción tiene por objeto gestionar valores de datos continuos mediante la construcción de un modelo (Al. Barrak, 2015).

Predicción / pronóstico de un modelo se refiere a su capacidad para predecir la mejor respuesta o salida más cercana a la realidad, sobre la base de los datos de entrada.

2.4.2 Modelo de predicción de la deserción universitaria

El modelo de predicción de la deserción universitaria es una representación abstracta, que emplea estructuras para predecir con aceptable grado de precisión la probabilidad de que un estudiante universitario abandone sus estudios en el primer año de universidad, dados los valores de sus atributos en un momento dado.

La finalidad del modelo de predicción de la deserción universitaria es predecir si un estudiante tiene riesgo de desertar de la universidad como su fin inmediato, el modelo de predicción de debe ser eficaz en la identificación temprana de estudiantes con riesgo de abandonar los estudios universitarios (Lara. et all., 2014). Debe operar como un sistema de alerta en la detección de estudiantes riesgo y se convertirá en una herramienta esencial para la aplicación de políticas que permitan reducir las tasas de deserción en las universidades (Huang, 2013).

Para la construcción del modelo de predicción de deserción se debe considerar una adecuada selección de las variables de ingreso y de técnicas

de procesamiento de datos adecuadas para la obtención de resultados predictivos exitosos.

Además, un modelo de predicción puede ayudar a los profesores a desarrollar estrategias de intervención eficaz para reducir la tasa de deserción de estudiantes cursos o programas de pregrado.

CAPÍTULO III: ESTADO DEL ARTE

3.1 Introducción

En la actualidad, existe un creciente interés por investigar el tema de la deserción estudiantil en las universidades alrededor del mundo (Márquez, Morales & Soto, 2013) una de las principales preocupaciones es su alta tasa (Yukselturk, Ozekes & Turel, 2014). La deserción afecta negativamente a las Instituciones de Educación Superior en la reducción de la matrícula y el incumplimiento de los objetivos institucionales (Lin, 2012). Como consecuencia, estudiantes, universidades y gobiernos se ven afectados, tanto en aspectos económicos y sociales. Además, el abandono estudiantil se convierte en un tema crítico cuando los administradores universitarios no poseen las herramientas oportunas para identificar a los estudiantes que están en riesgo de abandonar las instituciones educativas, de este modo, se disminuye el accionar de medidas correctivas necesarias que permitan la retención de los estudiantes en las universidades. (Hu, Lo & Shih, 2014)

De igual manera, la predicción temprana de abandono escolar se ha convertido en un desafío importante, para la identificación de factores que contribuyen a que este fenómeno no ocurra constantemente (Lye, Hassan, Goh, Law & Ismail, 2010). Una de las posibles razones por las cuales todavía existen altas tasas de abandonos universitarios puede estar asociada con el hecho de que la mayoría de los modelos de predicción aplicados para resolver este problema son difíciles de interpretar. (Wray, Barrett, Aspland & Gardiner, 2012)

Se ha realizado un esfuerzo importante para cerrar la brecha de la deserción universitaria y disminuir sus tasas. Sin embargo, esto no ha sido suficiente

(Djulovic & Li, 2013); según la Organización para la Cooperación y el Desarrollo (OCDE), en el año 2016, la tasa de deserción en Europa se encontró entre el 30 y el 50 %, mientras que en Estados Unidos fue del 37% (Rodríguez, Vindas, 2005). En algunos países latinoamericanos como Colombia, la deserción superó el 40 %, mientras que en Brasil alcanzó aproximadamente el 54 %. En Costa Rica, la tasa de deserción alcanzó el 50 % (ONU, 2013), problemas que se evidencian con mayor frecuencia en las universidades públicas.

Una de las medidas para tratar el problema de la deserción universitaria se basa en predecir sus tasas; para ello se utiliza la minería de datos que consiste en el desarrollo de métodos para extraer patrones en grandes colecciones de información y obtener conocimiento (Yukselturk, Ozekes & Turel, 2014). El proceso de predicción a través de este enfoque se utiliza ampliamente para estudiar la deserción, debido a que permite predecir con un aceptable grado de significancia. En general, este proceso sigue cuatro etapas que van desde el preprocesamiento de datos hasta la evaluación de los resultados (ver figura 1).

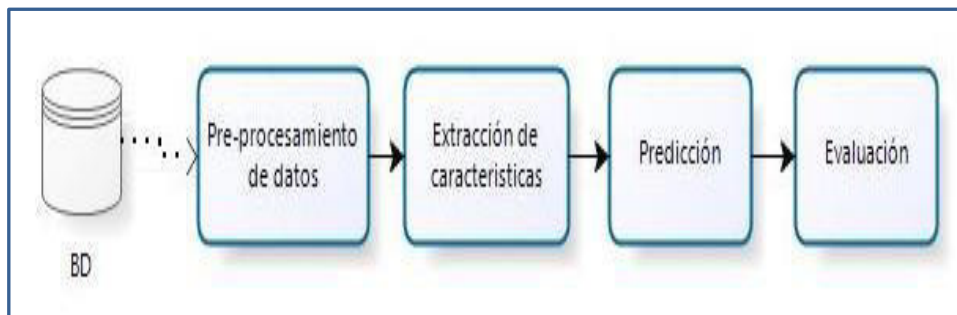


Figura 2. Proceso de Minería de Datos para la Predicción de la Deserción Estudiantil Universitaria.

Fuente: Adaptado de Sagodiah, Muniandy, Heng y Remendran (2015).

Por otro lado, se identificaron estudios previos de revisión de literatura Romero & Ventura (2010), Papamitsiou & Economices (2014). El primer estudio presenta un resumen de las tareas más comunes en el ambiente educativo que han sido resueltas a través de técnicas de minería de datos relacionadas con la educación tradicional, educación E-learning, sistemas

para el manejo del aprendizaje, sistemas de tutorías inteligentes, sistemas adaptativos educacionales, cuestionarios y test de contenidos. De igual manera, Papamitsiou & Economices (2014) presentan un resumen de evidencias empíricas sobre investigaciones relacionadas con la analítica del aprendizaje, la minería de datos educacional y su impacto en el aprendizaje adaptativo. Además, los autores realizan una clasificación y análisis de los documentos de acuerdo a los casos de estudio sobre entornos de aprendizaje, métodos de extracción de datos y objetivos de investigación. En cada uno de los entornos educativos analizados, se consideraron aspectos de modelamiento de estudiantes, predicción del rendimiento, predicción de la deserción y la retención, mejora de los servicios de evaluación, retroalimentación y recomendación de recursos.

Si bien los documentos analizados son de gran importancia en el tema objeto de estudio, sin embargo, en general, los estudios no consideran el entorno de aplicación, los factores de predicción establecidos como variables de ingreso a los modelos predictivos y la precisión de las técnicas de minería de datos aplicadas.

Por esta razón, este trabajo tiene como objetivo responder a la siguiente pregunta: ¿Cuáles son los aspectos considerados en la predicción de la deserción estudiantil universitaria, a través de la minería de datos? Para responder esta pregunta de investigación, se han planteado cinco preguntas secundarias, las mismas que se presentan en la sección 2.1, y que permitieron organizar, analizar y discutir los estudios potenciales relacionados al tema objeto de estudio.

3.2 Metodología de Investigación

Para la revisión sistemática realizada, se consideró las metodologías aplicadas por Suhirman, Herawan, Chiroma & Zain (2014), así como también la metodología propuesta por Kitchenham (2004) y consta de tres etapas:

Planificación de la revisión: Identifica la necesidad de investigación y la determinación de un protocolo de revisión.

Conducción de la revisión: Fase de la ejecución del plan, se aplica el protocolo definido, así como los criterios de inclusión y exclusión.

Reporte de la revisión: Presenta los resultados estadísticos y se realiza el análisis del contenido de los documentos seleccionados.

3.1.1 Planificación de la Revisión:

Se plantearon cinco preguntas de investigación con la finalidad de poder determinar los aspectos desarrollados para predecir la deserción estudiantil universitaria.

Q1: ¿Cuáles son los factores usados para predecir la deserción universitaria?

Q2: ¿Cuáles son las técnicas utilizadas para el pre procesamiento de los datos?

Q3: ¿Cuáles son las técnicas utilizadas para la extracción de factores?

Q4: ¿Cuáles son las técnicas usadas para la predicción y su nivel de confiabilidad?

Q5: ¿Cuáles son las herramientas utilizadas?

Se revisaron artículos de conferencias y de journals con factor de impacto en los siguientes bancos: Scimago Journal Country Rank (SJR), Science Direct, ACM Digital Library, IEEE Explore, Springer, DOAJ, Taylor and Francis, Emerald, Proquest y Ebsco. Para la selección de los documentos, se aplicaron los criterios de inclusión y exclusión, que se presentan en la Tabla 3.

Se consideraron los criterios de búsqueda (“dropout student” OR “dropout out student” OR “dropping student” AND “data mining”) que fueron

aplicados al título, resumen y palabras claves, en el período de búsqueda comprendido entre enero de 2016 a junio de 2018.

Tabla 3
Criterios para selección de documentos

Inclusión	Exclusión
<ul style="list-style-type: none"> - Modelos para dar solución al problema de la deserción estudiantil universitaria - Identifica factores que inciden en la deserción en las universidades - Artículos que incluyen técnicas de preprocesamiento, extracción de factores y predicción basados en minería de datos - Aplican métricas para evaluar la calidad de los modelos predictivos - Responden a las preguntas de investigación 	<ul style="list-style-type: none"> - Documentos de predicción no relacionados a la deserción estudiantil universitaria como educación básica, secundaria y posgrado - Trabajos que no se basan en minería de datos - Trabajos que no presentan experimentación numérica. - Documentos que no se encuentran dentro del periodo de búsqueda establecido

3.2.2 Conducción de la Revisión

El proceso de búsqueda se realizó con base a las estrategias planteadas en la sección 2.1. Una vez seleccionado los documentos, se realizó una revisión de su contenido para determinar la relación respecto a los criterios de selección establecidos. El proceso de revisión sistemática de la literatura se presenta en la Figura 3.

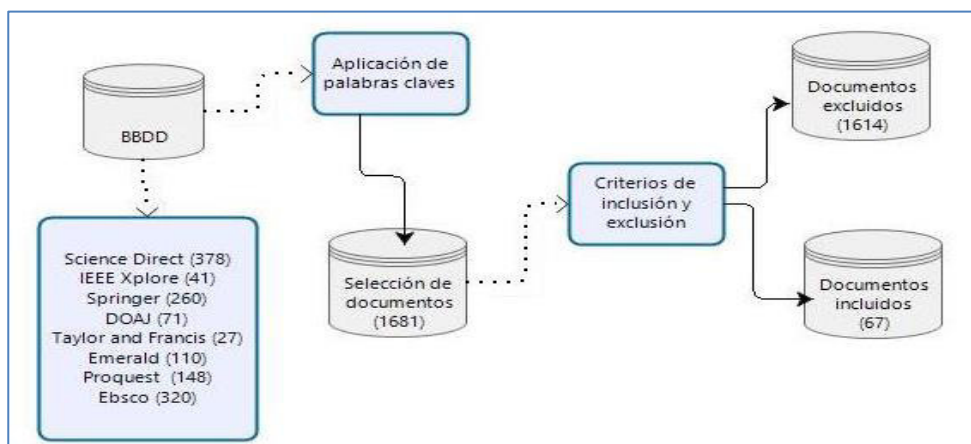


Figura 3. Proceso para la revisión sistemática de literatura

Fuente: Elaborado por el autor

3.3 Resultados

La tabla 4 resume el total de documentos encontrados y seleccionados por fuente de información, siendo Science Direct la principal fuente de información con un 40 % de estudios primarios seleccionados, mientras que ACM Digital Library y Emerald presentan un porcentaje del 1.49 % y 4.47 % respectivamente.

Tabla 4
Producción científica seleccionada

Fuente	Artículos Identificados	Artículos Seleccionados
Science Direct	378	27
ACM Digital Library	326	1
IEEE Xplore	41	10
Springer	260	6
DOAJ	71	5
Taylor and Francis	27	5
Emerald	110	3
Proquest	148	4
Ebsco	320	6
Total	1681	67

La figura 4 evidencia el incremento de investigaciones en el transcurso de los últimos doce años y el interés que existe por parte de los investigadores para resolver el problema de la predicción de la deserción estudiantil universitaria.

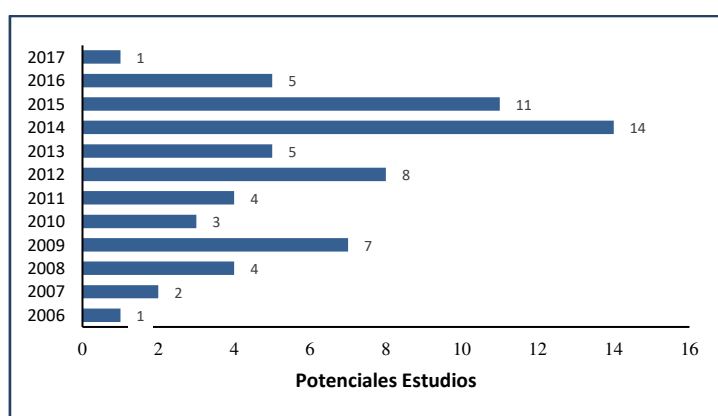


Figura 4. Tendencia temporal de publicaciones seleccionadas sobre deserción estudiantil universitaria.

Fuente: Elaborado por el autor

El 87 % de documentos primarios seleccionados proceden de journals (58 estudios de 67) y el 13 % corresponde a publicaciones en conferencias (9 estudios de 67) como se visualiza en la figura 5.

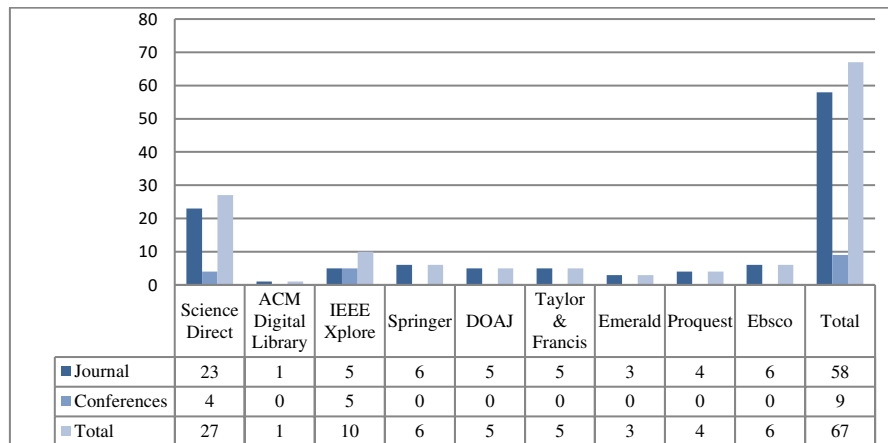


Figura 5. Publicaciones sobre predicción de la deserción universitaria por tipo de fuente.

Fuente: Elaborado por el autor

De los documentos seleccionados, se identificó 3 aspectos sobre la predicción de la deserción estudiantil universitaria: factores, técnicas y herramienta, que a continuación se precisan en el marco del presente estudio.

Factores de deserción: Son las razones por las cuales los estudiantes abandonan los estudios Girón (2014).

Técnicas de minería de datos: Tiene como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones y técnicas avanzadas de análisis de datos Ballesteros & Iñiguez (2018).

Herramientas de minería de datos: Se refiere al software que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y para predecir comportamientos futuros, Romero, Ventura, Pechenizkiy & Bayer (2010).

3.3.1 Q1: ¿Cuáles son los factores usados para predecir la deserción universitaria?

Se identificaron 112 factores usados para predecir la deserción estudiantil universitaria, los mismos que fueron clasificados siguiendo las 5 dimensiones (personales, académicos, económicos, sociales e institucionales).

Factores personales: Constituidos por características que determinan el comportamiento del alumno, tales como sentimientos, pensamientos o acciones; decisivos en el desarrollo de su entorno educativo. Se identificaron 31 factores en la categoría personales y responden aproximadamente al 28 % del total de factores identificados, los cuales se muestran en la tabla 5.

Tabla 5
Factores de la dimensión personal

Factor	Referencia
Ajuste	Sivakumar, Venkataraman & Selvaraj (2016)
Edad	Yukselturk, Ozekes & Turel (2014), Wray, Brett, Aspland & Gardiner (2012), Tan & Shao (2015), Levy (2007), Oeda & Hashimoto (2017), Park & Choi (2009), Houdhaugen (2009), Chen & Desjardins (2008), Herzog (2005), Tumen, Shulruf & Hattie (2008), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015), Willging & Johnson (2009), Arulampalam, Naylor & Smitg (2007), González, Heracleous & Winters (2012), Stratton, O'Toole & Wetzel (2008), Arulampalam, Naylor & Smith (2005), Di Pietro & Cutillo (2008), Melguizo, Sanchez & Velasco (2016)
Cambio de meta	Saranya & Rajeswari (2016), Sivakumar, Venkataraman & Selvaraj (2016)
Elección para cambiar el curso actual	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
País o ciudad de origen	Hoffait & Schyns (2017),Guarín, Guzmán & González (2015),Heredia, Amaya & Barrientos (2015)
Dependientes	Wray, Brett, Aspland & Gardiner (2012)
Discapacidad	Wray, Brett, Aspland & Gardiner (2012)
Dirección	Wray, Brett, Aspland & Gardiner (2012), Lam-ON & Boongoen (2014), Levy (2007),Di Pietro & Cutillo (2008), Sivakumar, Venkataraman & Selvaraj (2016), Tan & Shao (2015)
Apoyo de los padres	Arulampalam, Naylor & Smitg (2007)
Compromiso del estudiante	Houdhaugen (2009), Reschly & Christenson (2006),Willcoxson, Cotter & Joy (2011)

Tabla 5
Factores de la dimensión personal (Continuación)

Factor	Referencia
Etnia	Wray, Brrett, Aspland & Gardiner (2012), Thammasiri, Delen, Meesad & Kasap (2014), Tan & Shao (2015), Chen & Desjardins (2008), Herzog (2005), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Guarín, Guzmán & González (2015), Martinho, Nunes & Minussi (2013), Willging & Johnson (2009), Arulampalam, Naylor & Smitg (2007), González, Heracleous & Winters (2012), Oseguera & Rhee (2009)
Género	Yukselturk, Ozekes & Turel (2014), Wray, Brrett, Aspland & Gardiner (2012), Yasmin (2013), Tan & Shao (2015), Yi et al. (2015), Natek & Zwilling (2014), Lam-ON & Boongoen (2014), Paura & Arhipova (2014), Hoffait & Schyns (2017), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Levy (2007), Jadric, Garaca & Cukusic (2010), Park & Choi (2009), Houdhaugen (2009), Sultana, Khan & Abbas (2017), Chen & Desjardins (2008), Herzog (2005), Tumen, Shulruf & Hattie (2008), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Iepsen, Bercht & Reategui (2013), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015), Willging & Johnson (2009), Melguizo, Sanchez & Velasco (2016)
Problemas de salud	Saranya & Rajeswari (2016)
Nivel de interés	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Motivación intrínseca	Arifin (2016), Willcoxson, Cotter & Joy (2011)
Nivel de compromiso	Human_Vogel & Rabe (2015)
Vive en el campus	Herzog (2005)
Soledad	Alkan (2014)
Estado civil	Jia & Maloney (2015), Yasmin (2013), Jadric, Garaca & Cukusic (2010), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015), Stratton, O'Toole & Wetzel (2008)
Medida de persistencia del alumno	Duarte, Ramos, Goncalves (2014)
Pesimismo	Oeda & Hashimoto (2017), Iepsen, Bercht & Reategui (2013)
Lugar de residencia	Tan & Shao (2015), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Herzog (2005)
Autoeficacia	Hoffait & Schyns (2017), Willcoxson, Cotter & Joy (2011), Arifin (2016), Yukselturk, Ozekes & Turel (2014)
Satisfacción del estudiante	Saranya & Rajeswari (2016), Levy (2007), Duque (2014), Alkan (2014)
Fuente de matrícula	Tan & Shao (2015)
Implicación vocacional	Belo & Oliveira (2015)
Experiencia	Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)
Año de nacimiento	Natek & Zwilling (2014)

Para muchos autores, los factores de tipo personal son la principal causa que originan los abandonos en las carreras universitarias, la tabla 4 evidencia este hecho. Los factores edad y género son los factores con mayor frecuencia de uso para la predicción de la deserción. Esto se debe a que se los considera como factores internos de variabilidad, que resultan simples de definir y medir Yasmin (2013).

Factores académicos: Hacen referencia al desarrollo del alumno en su proceso formativo. Se identificaron 40 factores académicos que corresponden al 36 % del total de factores identificados, los cuales se presentan en la tabla 6.

Tabla 6

Factores de la dimensión académica usados para predecir la deserción universitaria.

Factor	Referencia
Absentismo	Sultana, Khan & Abbas (2017)
Habilidad académica	Di Pietro & Cutillo (2008), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012)
Sobrecarga académica	Di Pietro & Cutillo (2008)
Desempeño académico	Jia & Maloney (2015), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Melguizo, Sanchez & Velasco (2016)
Edad de ingreso	Al-Barrak & Al-Razgan (2015)
Resultado de la evaluación formativa	Tan & Shao (2015)
Mejor desempeño en el examen GPA, SAT	Lin (2012), Jia & Maloney (2015), Thammasiri, Delen, Meesad & Kasap (2014), Duarte, Ramos, Goncalves (2014), Aulck, Velagapudi, Blumenstock & West (2016), Lam-ON & Boongoen (2014), Levy (2007), Huang & Fang (2013), Koonce & Hening (2009), Chen (2012), Reschly & Christenson (2006), Alkan (2014), Chen & Desjardins (2008), Herzog (2005), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Guarín, Guzmán & González (2015), Willging & Johnson (2009), Oseguera & Rhee (2009), Human-Vogel & Rabe (2015)
Cohorte	Wray, Brrett, Aspland & Gardiner (2012), Yi et al. (2015), Willging & Johnson (2009), González, Heracleous & Winters (2012)
Participación curricular	Martinho, Nunes & Minussi (2013)
Título	Jia & Maloney (2015), Wray, Brrett, Aspland & Gardiner (2012), Natek & Zwilling (2014), Chen (2012), Jadric, Garaca & Cukusic (2010), Chen & Desjardins (2008)
Aspiración de grado	Oseguera & Rhee (2009)
Duración del programa de grado	Arulampalam, Naylor & Smith (2005)
Intensión de desertar	Alkan (2014)

Tabla 6

Factores de la dimensión académica usados para predecir la deserción universitaria (Continuación)

Factor	Referencia
Meta educativa	Houdhaugen (2009)
Idioma Inglés	Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)
Inscrito en otra institución	Saranya & Rajeswari (2016)
Calificaciones de ingreso	Wray, Brrett, Aspland & Gardiner (2012), Duarte, Ramos, Goncalves (2014)
Experiencia	Yukselturk, Ozekes & Turel (2014), Wray, Brrett, Aspland & Gardiner (2012)
Examen final	Tan & Shao (2015), Cukusic, Garaca & Jadric (2014), Xing, Chen, Stein & Marcinkowski (2016), Huang & Fang (2013), Sultana, Khan & Abbas (2017), Al-Barrak & Al-Razgan (2015)
Examen parcial	Al-Barrak & Al-Razgan (2015)
Créditos del primer semestre	Herzog (2005)
Motivo de elección de la carrera	Houdhaugen (2009)
Cuestionario	Al-Barrak & Al-Razgan (2015)
Participa en actividades extracurriculares	Saranya & Rajeswari (2016), Sivakumar, Venkataraman & Selvaraj (2016)
Calificaciones de la secundaria	Jadric, Garaca & Cukusic (2010), Sultana, Khan & Abbas (2017), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Progreso en los estudios	Wray, Brrett, Aspland & Gardiner (2012)
Preparación	Yukselturk, Ozekes & Turel (2014)
Créditos reconocidos	Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Tan & Shao (2015), Heredia, Amaya & Barrientos (2015)
Uso de recursos	Belo & Oliveira (2015)
Satisfacción del curso	Sivakumar, Venkataraman & Selvaraj (2016)
Puntuación de integración académica	Chen (2012), Chen & Desjardins (2008), Al-Barrak & Al-Razgan (2015)
Calificaciones	Paura & Arhipova (2014), Hoffait & Schyns (2017), Heredia, Amaya & Barrientos (2015), Arulampalam, Naylor & Smitg (2007), Al-Barrak & Al-Razgan (2015), Melguizo, Sanchez & Velasco (2016)
Autoevaluación	Sultana, Khan & Abbas (2017), Martinho, Nunes & Minussi (2013)
Estado de matrícula de estudiantes	Sultana, Khan & Abbas (2017), Arulampalam, Naylor & Smith (2005), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Centro de estudios	Tan & Shao (2015), Martinho, Nunes & Minussi (2013), Arulampalam, Naylor & Smitg (2007), Stratton, O'Toole & Wetzal (2008)
Nivel de estudios	Tan & Shao (2015), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Park & Choi (2009), Herzog (2005)
Jornada de estudios	Martinho, Nunes & Minussi (2013)
Tasa de éxito	Jia & Maloney (2015), Tan & Shao (2015)
Apoyo para el aprendizaje	Willcoxson, Cotter & Joy (2011)
Total de cursos reprobados	Tan & Shao (2015)

Los resultados del análisis de estos factores indican que la prueba de ingreso a la universidad es el factor con mayor frecuencia de uso en las investigaciones analizadas. Se puede señalar que el proceso de aprendizaje en la universidad guarda una estrecha relación con los niveles de estudio precedentes, que inciden en los logros educativos posteriores Nistor & Neubauer (2010).

De igual manera, la calificación que obtiene un estudiante en el examen de ingreso a la universidad puede constituirse en un indicador para explicar el éxito o fracaso en la trayectoria académica universitaria Jia & Maloney (2015). En este sentido, varios estudios han analizado la validez predictiva de este factor, considerándolo como un predictor de características cognitivas y actitudinales importantes para el éxito de los estudiantes en la universidad Sangodiah, Beleya, Muniandy & Ramendran (2015).

Factores económicos: Se relacionan con las condiciones que tienen los estudiantes para satisfacer las necesidades económicas que se presentan mientras cursan un programa académico. En esta dimensión, se identificaron 15 factores que inciden en la deserción y corresponden aproximadamente al 13 % del total de factores analizados, los cuales se presentan en la Tabla 7.

Tabla 7

Factores de la dimensión económicos usados para predecir la deserción universitaria.

Factor	Referencia
Beca	Lin (2012),Thammasiri, Delen, Meesad & Kasap (2014), Hoffait & Schyns (2017)
Pertenece al cuartil de pobreza	Yasmin (2013)
Trabaja en el campus	Herzog (2005)
Dependencia	Arulampalam, Naylor & Smitg (2007)
Préstamo estudiantil	Thammasiri, Delen, Meesad & Kasap (2014)
Ingresos familiares	Martinho, Nunes & Minussi (2013)
Trabajo del padre	Jia & Maloney (2015)
Preocupaciones financieras	Oseguera & Rhee (2009),Willcoxson, Cotter & Joy (2011), Lin (2012)
Inversión	Human-Vogel & Rabe (2015)
Ingreso bruto de los padres	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)

Tabla 7

Factores de la dimensión económicos usados para predecir la deserción universitaria (Continuación).

Factor	Referencia
Préstamos recibidos	Lin (2012), Thammasiri, Delen, Meesad & Kasap (2014), Chen (2012)
Estado de empleo del estudiante	Sultana, Khan & Abbas (2017)
Tasa de estudiantes	Arulampalam, Naylor & Smith (2005)
Tipo de ayuda financiera	Thammasiri, Delen, Meesad & Kasap (2014), Paura & Arhipova (2014), Chen (2012), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)

Los factores de la dimensión económica aluden a las comodidades materiales y a la capacidad de los padres para destinar más y mejores recursos para el desempeño académico de los hijos, los cuales inciden significativamente en el rendimiento académico, Arulampalam, Naylor & Smith (2007).

Factores sociales: Son aspectos que afectan a los estudiantes en su conjunto, determinados por el lugar y el espacio en el que se encuentran, se presentan en la tabla 8.

Tabla 8

Factores de la dimensión social usados para predecir la deserción estudiantil universitaria.

Factor	Referencia
Alojamiento en el campus	Arulampalam, Naylor & Smitg (2007), Arulampalam, Naylor & Smith (2005), Oseguera & Rhee (2009)
Categoría marginada o sección vulnerable de la sociedad	Yasmin (2013), González, Heracleous & Winters (2012)
Estado de la Universidad	Levy (2007)
Soporte comunitario	Sultana, Khan & Abbas (2017)
Estado del empleo	Yasmin (2013), Natek & Zwilling (2014), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Martinho, Nunes & Minussi (2013)
Problemas familiares	Sivakumar, Venkataraman & Selvaraj (2016)
Composición familiar	Sivakumar, Venkataraman & Selvaraj (2016)
Nivel de educación del padre	Jia & Maloney (2015), Jadric, Garaca & Cukusic (2010), Houdhaugen (2009), Chen & Desjardins (2008), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015)
Indicador de vivienda	Jadric, Garaca & Cukusic (2010)
Nivel de implicación de las redes sociales	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)

Tabla 8

Factores de la dimensión social usados para predecir la deserción estudiantil universitaria (Continuación)

Factor	Referencia
Medio de transporte	Martinho, Nunes & Minussi (2013)
Migrante	Yi, et al. (2015), Houdhaugen (2009)
Nivel económico de la madre	Yi, et al. (2015)
Madre migrante	Yi, et al. (2015)
Nivel de educación de la madre	Jia & Maloney (2015), Jadric, Garaca & Cukusic (2010), Houdhaugen (2009), Sultana, Khan & Abbas (2017), Chen & Desjardins (2008), Martinho, Nunes & Minussi (2013), Melguizo, Sanchez & Velasco (2016)
Ocupación de la madre, padre	Heredia, Amaya & Barrientos (2015), Willging & Johnson (2009)
Ocupación de los parientes	Saranya & Rajeswari (2016)
Estatus político	Tan & Shao (2015)
Estatus social	Jadric, Garaca & Cukusic (2010), Park & Choi (2009), Heredia, Amaya & Barrientos (2015), Arulampalam, Naylor & Smitg (2007), Belo & Oliveira (2015), Arulampalam, Naylor & Smith (2005)
Stress	Saranya & Rajeswari (2016)
El estudiante consume drogas	Arulampalam, Naylor & Smitg (2007)
Usa las instalaciones recreativas	Herzog (2005)

Por otro lado, la dimensión social enfoca la importancia de la interacción del alumno con su entorno social; dicha interacción es el marco en el que se constituye el sujeto en su relación con la norma académica, la institución y los hábitos de estudio. (Xing, Chen, Stein & Marcinkowski, 2016)

Factores institucionales: Los factores que corresponden a esta categoría están relacionados a las características estructurales y funcionales de una institución, los cuales se presentan en la tabla 9 y corresponden aproximadamente al 3.53 % del total de factores analizados.

Tabla 9

Factores de la dimensión institucional usados para predecir la deserción estudiantil universitaria.

Factor	Referencia
Ambiente del campus	Sivakumar, Venkataraman & Selvaraj (2016)
Tipo de escuela secundaria	Guarín, Guzmán & González (2015)
Involucramiento institucional	Belo & Oliveira (2015)
Infraestructura de la universidad	Sivakumar, Venkataraman & Selvaraj (2016)

3.3.2 Q2: ¿Cuáles son las técnicas utilizadas para el preprocesamiento de datos?

En la etapa de preprocesamiento, se identificaron 11 técnicas que se presentan en la tabla 10. Esta etapa permite el manejo de anomalías, correcciones de valores atípicos y valores faltantes (Wray, Barrett, Aspland & Gardiner, 2012). Su finalidad es mejorar las propiedades de las variables y resolver las anomalías de los datos para optimizar el proceso de búsqueda de los algoritmos de minería de datos (Vargas, Pinzón & Patiño, 2008), y constan de tres actividades: integración, limpieza y transformación de la información.

Todos los estudios que involucran la etapa de preprocesamiento de datos se concentran en la actividad de transformación de los datos, siendo las técnicas de normalización y discretización las más usadas. Sin embargo, las actividades de limpieza e integración también son importantes, tal como lo señala Lara (2014) y Tan (2015), una entrada de variables erróneas en el proceso de minería de datos puede afectar negativamente la precisión de las técnicas aplicadas para predecir la deserción.

Tabla 10
Técnicas para el preprocesamiento de datos

Técnica	Referencia
Análisis multivariante de la varianza	Reschly & Christenson (2006)
Bagging	Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Chen (2012)
Discretización	Wray, Brrett, Aspland & Gardiner (2012), Saranya & Rajeswari (2016), Tan & Shao (2015), Badr, Algobail, Almutairi & Almutery (2016),[Al-Barrak & Al-Razgan (2015)
Filtrado basado en atributos	Saranya & Rajeswari (2016)
Imputación de datos	Aulck, Velagapudi, Blumenstock & West (2016)
Multiple imputation	Koonce & Hening (2009),Chen (2012)
Normalización	Lara, Lizcano, Martinez, Pazos & Riera, Saranya & Rajeswari (2016), Xing, Chen, Stein & Marcinkowski (2016), Badr, Algobail, Almutairi & Almutery (2016), Huang & Fang (2013), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Sivakumar, Venkataraman & Selvaraj (2016)
Sobre muestreo	Yasmin (2013)
Muestro aleatorio simple	Jadric, Garaca & Cukusic (2010), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Imputación simple	Koonce & Hening (2009)

3.3.3 Q3: ¿Cuáles son las técnicas utilizadas para la extracción de factores?

Se identificó 10 técnicas aplicadas para la extracción de factores, el objetivo de la aplicación de estas técnicas es seleccionar los factores más relevantes

que sirven como variables de ingreso a los modelos de predicción de la deserción.

Tabla 11
Técnicas para la extracción de factores

Técnica	Referencia
Análisis de la varianza	Yasmin (2013), Cukusic, Garaca & Jadric (2014), Reschly & Christenson (2006), Levy (2007), Hershkovitz & Nachmias (2011)
Estadística descriptiva	Jia & Maloney (2015), Yasmin (2013), Duarte, Ramos, Goncalves (2014), Cukusic, Garaca & Jadric (2014), Paura & Arhipova (2014), Nistor & Neubauer (2010), Oeda & Hashimoto (2017), Huang & Fang (2013), Chen (2012), Park & Choi (2009), Duque (2014), Alkan (2014), Chen & Desjardins (2008), Herzog (2005), Zacharis (2015), Tumen, Shulruf & Hattie (2008), Arulampalam, Naylor & Smitg (2007), Martinho, Nunes & Minussi (2013), Willging & Johnson (2009), González, Heracleous & Winters (2012), Belo & Oliveira (2015), Stratton, O'Toole & Wetzel (2008), Arulampalam, Naylor & Smith (2005), Di Pietro & Cutillo (2008), Hershkovitz & Nachmias (2011), Oseguera & Rhee (2009), Willcoxson, Cotter & Joy (2011), Human-Vogel & Rabe (2015), Melguizo, Sanchez & Velasco (2016), Arbiv & Meiran (2015), Sivakumar, Venkataraman & Selvaraj (2016), Arifin (2016)
Algoritmo de extracción de características	Li, Gao, Xiong, Wen & Wu (2016)
Algoritmo genético	Tan & Shao (2015)
Hosmer and Lemeshow	Jia & Maloney (2015), Duarte, Ramos, Goncalves (2014)
Locality Preserving Projection (LPP)	Xing, Chen, Stein & Marcinkowski (2016)
Máxima verosimilitud	Koonce & Hening (2009)
Neighborhood Preserving Embedding (NPE)	Xing, Chen, Stein & Marcinkowski (2016)
Análisis de componentes principales (PCA)	Saranya & Rajeswari (2016), Xing, Chen, Stein & Marcinkowski (2016), Natek & Zwilling (2014), Lam-ON & Boongoen (2014), Houdhaugen (2009)
Prueba de kaiser Meyer Olkin (KMO)	Jia & Maloney (2015), Elffers (2013)
U Mann Whitney	Belo & Oliveira (2015)

Aproximadamente el 55 % (23 de 42 estudios) utilizaron la estadística descriptiva debido a que permite obtener características de dispersión, localización y formas de distribución de las variables (Chen, 2012). Además, es frecuentemente utilizada para identificar patrones sobre características de los estudiantes y los comportamientos asociados con la deserción (Duarte, 2014). 14 de 23 trabajos que se encuentran en la clasificación de la estadística

descriptiva se orienta a correlación de variables y aplica este tipo de análisis para evaluar la asociación y relación de datos cuantitativos en términos de direccionalidad, mediante coeficientes de correlación. (Zacharis, 2015)

Por otro lado, el 12 % (5 de 42 estudios) aplican la técnica de Análisis Principal de Componentes, para reducir la dimensionalidad de las variables observadas a un número menor de variables hipotéticas, para lo cual se establece grupos de variables que se correlacionan con otras, transformándolas en factores independientes que ingresan a los modelos predictivos de deserción. (Hovdhaugen, 2009)

3.3.4 Q4: ¿Cuáles son las técnicas utilizadas para la predicción de la deserción universitaria?

Se identificaron 14 técnicas de minería de datos que han sido clasificadas en: inteligencia artificial y métodos estadísticos, que se presentan en las tablas 12 y 13, respectivamente. Estas técnicas utilizan conocimiento para representar generalizaciones que pueden ser comprendidas y modificadas por los usuarios (Jadric, 2010) y son utilizadas para el descubrimiento de patrones y la generación de modelos predictivos. (López, 2004)

Tabla 12
Técnicas de inteligencia artificial

Id	Técnica	Referencia
IA1	Neural network classifier	Yukselturk, Ozekes & Turel (2014), Hu, Lo & Shih (2014), Lara, Lizcano, Martinez, Pazos & Riera, Thammasiri, Delen, Meesad & Kasap (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Tan & Shao (2015), Xing, Chen, Stein & Marcinkowski (2016), Lam-ON & Boongoen (2014), Hoffait & Schyns (2017), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Jadric, Garaca & Cukusic (2010), Dewan, Lin & Wen (2015), Sultana, Khan & Abbas (2017), Martinho, Nunes & Minussi (2013)
IA2	Support vector machine	Yukselturk, Ozekes & Turel (2014), Lara, Lizcano, Martinez, Pazos & Riera, Thammasiri, Delen, Meesad & Kasap (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Aulck, Velagapudi, Blumenstock & West (2016), Li, Gao, Xiong, Wen & Wu (2016), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Dewan, Lin & Wen (2015), Fei & Yeung (2015), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Liang, Yang, Wu, Li & Zhen (2016)

Tabla 12
Técnicas de inteligencia artificial (Continuación)

Id	Técnica	Referencia
IA3	Decision tree classifier	Yukselturk, Ozekes & Turel (2014), Hu, Lo & Shih (2014), Thammasiri, Delen, Meesad & Kasap (2014), Yasmin (2013), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Tan & Shao (2015), Aulck, Velagapudi, Blumenstock & West (2016), Xing, Chen, Stein & Marcinkowski (2016), Natek & Zwilling (2014), Lam-ON & Boongoen (2014), Li, Gao, Xiong, Wen & Wu (2016), Hoffait & Schyns (2017), Jadric, Garaca & Cukusic (2010), Pal (2012) Fei & Yeung (2015), Sultana, Khan & Abbas (2017), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Al-Barrak & Al-Razgan (2015), Guarín, Guzmán & González (2015), Heredia, Amaya & Barrientos (2015), Hershkovitz & Nachmias (2011), Liang, Yang, Wu, Li & Zhen (2016), Sivakumar, Venkataraman & Selvaraj (2016)
IA4	A priori algorithm	Aziz, Idris, Hassan, Jusoh & Emran (2018)
IA5	K-Nearest neighbor classifier	Yukselturk, Ozekes & Turel (2014), Tan & Shao (2015)
IA6	Radial basis function neighbor classifier	Hoffait & Schyns (2017), Dewan, Lin & Wen (2015)
IA7	Naive bayes	Yukselturk, Ozekes & Turel (2014), Hu, Lo & Shih (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Tan & Shao (2015), Xing, Chen, Stein & Marcinkowski (2016), Lam-ON & Boongoen (2014), Li, Gao, Xiong, Wen & Wu (2016), Sultana, Khan & Abbas (2017), Al-Barrak & Al-Razgan (2015), Guarín, Guzmán & González (2015)
IA8	Classification association rules mining	Badr, Algobail, Almutairi & Almutery (2016)
IA9	Fuzzy inference	Saranya & Rajeswari (2016)
IA10	Rule induction	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)

Aproximadamente, el 79 % (22 de 28 estudios) del total de artículos analizados usan Decision Tree Classifier. Para Yasmin (2013) y Heredia, Amaya & Barrientos (2015), esta técnica es utilizada por su flexibilidad para tratar datos de naturaleza numérica y categórica, transformaciones de las variables explicativas, factibilidad de interpretación de resultados y porque presenta mejores tasas de precisión. Pal (2012) y Sivakumar (2016) mencionan que el algoritmo ID3 (Decisión tree classifier) es eficaz cuando se

trata de clasificar datos de registros de históricos de estudiantes, así, presenta mayor sensibilidad que otros algoritmos.

Neural Networks Classifier y Support Vector Machine presentan la segunda mayor frecuencia de uso, esto se puede explicar debido a que estos enfoques de minería de datos son considerados como herramientas poderosas para resolver problemas de clasificación (Liang, Jiajun, *et.al.*, 2016) y son frecuentemente utilizadas por su simplicidad y fácil entendimiento. (Sivakumar, Subitha, Sivakumar & Rajalakshmi, 2016)

Por otro lado, se identificaron 4 técnicas estadísticas que corresponden al 36.3 % (4 de 13 técnicas) del total analizadas, resultados que se presentan en la tabla 13.

Tabla 13
Técnicas estadísticas

Id	Técnica	Referencia
E1	Regresión logística	Wray, Brett, Aspland & Gardiner (2012), Thammasiri, Delen, Meesad & Kasap (2014), Saranya & Rajeswari (2016), Duarte, Ramos, Goncalves (2014), Aulck, Velagapudi, Blumenstock & West (2016), Li, Gao, Xiong, Wen & Wu (2016), Hoffait & Schyns (2017), Fei & Yeung (2015), Houdhaugen (2009), Reschly & Christenson (2006), Sultana, Khan & Abbas (2017), Chen & Desjardins (2008), Herzog (2005), Tumen, Shulruf & Hattie (2008), Arulampalam, Naylor & Smitg (2007), Stratton, O'Toole & Wetzel (2008), Arulampalam, Naylor & Smith (2005), Di Pietro & Cutillo (2008), Liang, Yang, Wu, Li & Zhen (2016), Arbiv & Meiran (2015), Arifin (2016)
E2	Regresión lineal	Elffers (2013), Cukusic, Garaca & Jadric (2014), Yi et al. (2015), Paura & Arhipova (2014), Oeda & Hashimoto (2017), Koonce & Hening (2009), Chen (2012), Jadric, Garaca & Cukusic (2010), Park & Choi (2009), Duque (2014), Alkan (2014), Zacharis (2015), Willging & Johnson (2009), González, Heracleous & Winters (2012), Belo & Oliveira (2015), Oseguera & Rhee (2009), Willcoxson, Cotter & Joy (2011), Human-Vogel & Rabe (2015), Melguizo, Sanchez & Velasco (2016)
E3	Análisis discriminante	Nistor & Neubauer (2010)
E4	Modelo Probit	Jia & Maloney (2015)

Aproximadamente el 54 % (21 de 39 estudios) aplicaron regresión logística y lineal. El uso de estas técnicas se debe a que son consideradas como una

excelente opción para el pronóstico de los estudiantes con riesgo de desertar (Cukusic, Garaca & Jadric, 2014) y son frecuentemente utilizadas para realizar clasificaciones basadas en características de los datos, por su flexibilidad para el uso de variables predictoras categóricas y continuas. (Herzog, 2005)

Con respecto a los estudios que tuvieron como propósito la identificación temprana de estudiantes en riesgo de deserción usando métodos estadísticos, aproximadamente, el 26 % (10 estudios de 39) toman como línea base la predicción de la deserción en el primer año de estudios, mientras que el 74 % (29 de 39 estudios) analizan la deserción a partir del segundo año de la universidad y en cada año de estudios.

Además, el 71 % (28 de 33 estudios) fueron realizados en entornos de formación universitaria en modalidad de estudios presencial, mientras que el 29% (5 de 33 estudios) fueron realizados en educación abierta en línea.

Por otro lado, con relación a la precisión de las técnicas de minería de datos, los autores han considerado métricas como: sensibilidad, especificidad y precisión. La métrica usada para este propósito es la precisión, que es determinado por el ratio de verdaderos positivos (TP) y verdaderos negativos (TN) entre el total de registros, como se formula en (1).

$$TP = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (1).$$

Donde FP corresponde a la cantidad de falsos positivos y FN a la cantidad de falsos negativos.

Las tablas 14 y 15 reportan el nivel de precisión de las técnicas de minería de datos que alcanzaron un ratio mayor al 60 % y tienen un data set compuesto por un número mayor a 100 estudiantes.

Tabla 14
Precisión de las técnicas estadísticas

Técnica	Conjunto de datos	Precisión (%)	Referencia
Regresión logística	237	71,80	Stratton, O'Toole & Wetzel (2008)
	6733	56,60	Chen & Desjardins (2008)
	293	85,50	Duarte, Ramos, Goncalves (2014)
	1064	85,80	Reschly & Christenson (2006)
Regresión lineal	588	87,80	Park & Choi (2009)
	37006	69,10	Oseguera & Rhee (2009)
	144	81,00	Oeda & Hashimoto (2017)
	134	81,30	Zacharis (2015)
Análisis discriminante	209	78,20	Nistor & Neubauer (2010)

Los resultados evidencian que la técnica regresión lineal presenta la mayor tasa de precisión con un 87 %. Sin embargo, estos resultados no se pueden generalizar, ya que dependen de la precisión de las técnicas, de las variables y del data set usado. Por otro lado, la tabla 15 presenta el porcentaje de precisión de las técnicas de inteligencia artificial usadas para predecir la deserción en las universidades.

Tabla 15
Precisión de las técnicas de inteligencia artificial

Conjunto de datos	Clasificador	Precisión (%)	Referencia
200	Feed forward neural network (FFNN)	82	Lara, Lizcano, Martínez, Pazos & Riera
	Support Vector Machine (SVM)	50	
	Probabilistic ensemble PESFAM	62	
	SEDM	94	
193	Feed forward neural network (FFNN)	84	Lykourentzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)
	Support Vector Machine (SVM)	83	
	Probabilistic ensemble simplified fuzzy ARTMAP	97	
170	Naive Bayes	81	Al-Barrak & Al-Razgan (2015)
	Jrip rule	55	
	J48	70	

Tabla 15
Precisión de las técnicas de inteligencia artificial (Continuación)

Conjunto de datos	Clasificador	Precisión (%)	Referencia
240	ID3	92,50	Sivakumar, Venkataraman & Selvaraj (2016)
	ID3 (Renyi)	97,50	
150	Support vector machine	89,84	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
	Decision tree	86,32	
	Rule induction	81,98	
3200	Naive Bayes	85	Guarín, Guzmán & González (2015)
	Decision Tree	57	
6043	Logistic regressions	32	Hoffait & Schyns (2017)
	Artificial neural networks	62	
	Decision trees and random forest	63	
62375	Artificial Neural Network (ANN)	84	Tan & Shao (2015)
	Decision Tree (DT)	82	
	Bayesian Networks (BNs)	76	
300	C4.5	98	Hu, Lo & Shih (2014)
	CART	97	
	Logistic Regresion	86	
775	Excalibur (J48) CSM	80	Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012)
	SNA (PART) CSM	92	
3617	General Bayesian Network (GBN)	89	Xing, Chen, Stein & Marcinkowski (2016)
	C4.5	86	
21654	Artificial Neural Networks (ANN)	85	Thammasiri, Delen, Meesad & Kasap (2014)
	Support Vector Machine (SVM)	90	
	Decision trees (DT)	89	
	Logistic regression	80	
	Regresión Logística	84	Li, Gao, Xiong, Wen & Wu (2016)
	Naive Bayes	83	
	Support Vector Machine	67	
	Decision trees (DT)	83	
128	Decision trees (DT)	84	Sultana, Khan & Abbas (2017)
	Logistic regression	84	
	Naive Bayes	82	
	Neural Networks (NN)	82	
189	K-Nearest Neighbour (k-NN)	87	Yukselturk, Ozekes & Turel (2014)
	Decision Tree (DT)	79	
	Naive Bayes (NB)	76	
	Neural Networks (NN)	73	

Tabla 15
Precisión de las técnicas de inteligencia artificial (Continuación)

Conjunto de datos	Clasificador	Precisión (%)	Referencia
32538	Logistic regression	66	Aulck, Velagapudi, Blumenstock & West (2016)
	Random Forest	62	
	K-Nearest Neighbour (k-NN)	64	
	ID3	90,90	Pal (2012)
	C4.5	89,09	
	CART	86,06	
	ADT	87,27	
200	K-Nearest Neighbour (k-NN)	74	Dewan, Lin & Wen (2015)
	Radial Basis Function Network Classifier	70	
	Support Vector Machine	79	
	Support Vector Machine	65	Liang, Yang, Wu, Li & Zhen (2016)
	Logistic regression	65	
	Random Forest	86	
	Gradient Boosting Decision Tree	88	

Los resultados indican que las técnicas que presentan la mayor confiabilidad de precisión son Decisión Tree Classifier en la aplicación de los clasificadores C4.5 y CART, con un 97 y 98 % respectivamente. Esto puede deberse a que esta técnica genera reglas que son fáciles de interpretar y permite reducir la probabilidad de errores de origen en un problema complejo. (Suhirman, Herawan, Chiroma & Zain, 2014)

2.3.5 Q5: ¿Cuáles son las herramientas utilizadas?

Se identificó 4 herramientas en aquellos estudios que aplicaron técnicas de inteligencia artificial y 7 herramientas usadas por las técnicas estadísticas, las cuales se presentan en las tablas 16 y 17.

Los resultados evidencian que las herramientas usadas con mayor frecuencia por las técnicas de inteligencia artificial y las estadísticas son WEKA y SPSS Modeler, las cuales se utilizan por tener una gran variedad de algoritmos de aprendizaje automático para tareas de minería de datos (Badr, Algobail, Almutairi & Almutery, 2016). La aplicación de estas herramientas se debe a

su flexibilidad en el modelado predictivo, así como también por su interfaz gráfica que permite acceder fácilmente a sus funcionalidades. (Xing, Chen, Stein & Marcinkowski, 2016)

Tabla 16
Herramientas usadas por las técnicas de inteligencia artificial

Herramientas	IA1	IA2	IA3	IA7	IA9
WEKA	Hu, Lo & Shih (2014), Thammasiri, Delen, Meesad & Kasap (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Natek & Zwilling (2014), Lam-ON & Boongoen (2014), Sultana, Khan & Abbas (2017), Heredia, Amaya & Barrientos (2015)	Thammasiri, Delen, Meesad & Kasap (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Natek & Zwilling (2014), Sultana, Khan & Abbas (2017), Heredia, Amaya & Barrientos (2015)	Hu, Lo & Shih (2014), Thammasiri, Delen, Meesad & Kasap (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Lam-ON & Boongoen (2014), Pal (2012), Sultana, Khan & Abbas (2017), Al-Barrak & Al-Razgan (2015), Guarán, Guzmán & González (2015)	Hu, Lo & Shih (2014), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Lam-ON & Boongoen (2014), Al-Barrak & Al-Razgan (2015), Guarán, Guzmán & González (2015)	Saranya & Rajeswari (2016)
SPSS Modeler	Thammasiri, Delen, Meesad & Kasap (2014), Hershkovitz & Nachmias (2011)	Thammasiri, Delen, Meesad & Kasap (2014), Hershkovitz & Nachmias (2011)	Thammasiri, Delen, Meesad & Kasap (2014)		
Matlab	Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Jadric, Garaca & Cukusic (2010)	Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)	Jadric, Garaca & Cukusic (2010)		

Tabla 16
Herramientas usadas por las técnicas de inteligencia artificial
 (Continuación)

Herramientas	IA1	IA2	IA3	IA7	IA9
SAS Enterprise	Jadric, Garaca & Cukusic (2010)		Jadric, Garaca & Cukusic (2010)		
Rapid Miner		Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Guarín, Guzmán & González (2015)	Guarín, Guzmán & González (2015)	

Tabla 17
Herramientas usadas por las técnicas estadísticas

Herramientas	E1	E2	E3
WEKA	Saranya & Rajeswari (2016), Thammasiri, Delen, Meesad & Kasap (2014), Sultana, Khan & Abbas (2017)	Zacharis (2015)	
SPSS Modeler	Wray, Brrett, Aspland & Gardiner (2012), Thammasiri, Delen, Meesad & Kasap (2014), Herzog (2005)	Elffers (2013), Paura & Arhipova (2014), Oeda & Hashimoto (2017), Zacharis (2015)	
Matlab		Jadric, Garaca & Cukusic (2010)	Aziz, Idris, Hassan, Jusoh & Emran (2018)
R	Arifin (2016)		
ISFE SYSTEM	Cukusic, Garaca & Jadric (2014)		
SAS Enterprise		Jadric, Garaca & Cukusic (2010)	
Excel	Jia & Maloney (2015)		

3.4 Discusión

Esta sección discute el análisis de los trabajos relacionados a la predicción de la deserción de los estudiantes en las universidades. Sobre la base de 67 artículos seleccionados, se establecen cuatro aspectos que relacionan el entorno de la educación superior, la deserción y la minería de datos: línea base de análisis de deserción, factores, técnicas y herramientas aplicados para resolver el problema de los abandonos universitarios.

Respecto a la línea base de la predicción de la deserción, aproximadamente el 67 % (45 de 67 estudios) analizan el riesgo de abandono universitario en primer año de la universidad basado en los siguientes enfoques: acciones preventivas sistemáticas de abandono, control de las trayectorias académicas, estudios motivados por los trabajos de Tinto y por la disponibilidad de los datos en relación a las cohortes de ingreso a la universidad, como un marco de tiempo apropiado para modelar el comportamiento de los estudiantes en cada año de la carrera, las políticas de admisión, la heterogeneidad de la oferta académica en el contexto de las instituciones universitarias, necesidades académicas insatisfechas, probabilidades de abandono y expectativas por terminar la universidad versus a sus condiciones de bajos ingresos, entre otras. Si bien existe evidencia empírica que sustenta los trabajos de investigación, no se puede evidenciar documentos que permitan contestar la siguiente pregunta: ¿En qué porcentaje la predicción de la deserción universitaria ayudará a reducir sus tasas?

Por otro lado, en cuanto a los factores que inciden en la predicción de la deserción, el análisis de los documentos permitió evidenciar que el 90 % (60 estudios) presentaron factores personales, académicos, económicos, sociales e institucionales como variables de ingreso a los modelos de predicción. Aun cuando las estadísticas permiten evidenciar que los factores edad, género, etnia y prueba de ingreso a la universidad son las variables más predominantes, varios autores centran su atención en factores como el rendimiento académico, nivel de educación del padre y la madre, y

determinan la importancia de ser considerado en los modelos de alerta temprana. Pese a que el conglomerado de factores encontrados es amplio, existe mucha controversia sobre qué tipo de factores resulta más eficiente al momento de predecir la deserción en las universidades.

En el enfoque de las técnicas estadísticas, se evidencia que la técnica regresión logística es la más usada en los procesos experimentales, mientras que las técnicas de inteligencia artificial Decision Tree Classifier, Neural Networks Classifiers y Support Vector Machine fueron consideradas como herramientas eficaces para extraer conocimiento sobre las características de los estudiantes al momento de establecer los estudiantes en riesgo de desertar. Si bien las técnicas de inteligencia artificial ofrecen varias ventajas en el procesamiento de grandes bases de datos, las técnicas estadísticas fueron utilizadas con mayor frecuencia en un 58 % (39 de 67 estudios), esto indica que aun cuando existe un incremento del uso de técnicas de inteligencia artificial para predecir la deserción, las estadísticas aún mantienen la preferencia de uso.

En lo que corresponde a la precisión de las técnicas de minería de datos utilizadas, se evidencia que Decisión Tree classifier y Support Vector Machine tienen la mayor tasa de precisión del 98 % y 97,5 %, sin embargo, estas tasas pueden variar dependiendo de las variables de ingreso, del contexto educativo estudiado, del entorno de educación aplicado, y de los antecedentes de los estudios para las que fueron aplicadas. Por otro lado, se considera importante determinar si es suficiente con predecir la deserción o si se requieren incorporar estudios que establezcan estrategias para mitigar la deserción en las instituciones de educación superior.

3.5 Conclusiones

El trabajo presenta una revisión sistemática de literatura sobre los aspectos de minería de datos considerados para predecir la deserción en las universidades. Los resultados de las búsquedas realizadas mostraron 1681 estudios primarios

relacionados con el tema objeto de estudio, de los cuales se seleccionaron 67 documentos de acuerdo a los criterios de inclusión y exclusión establecidos. Los resultados que arrojó el análisis de los documentos permitió identificar tres aspectos de minería de datos relacionados con la predicción de la deserción estudiantil universitaria, estos aspectos son los siguientes: factores, técnicas y herramientas, así, aproximadamente, el 67% de los documentos revisados cubrieron los tres aspectos antes mencionados.

Respecto a los factores que inciden en la predicción de la deserción, fueron clasificados en cinco dimensiones: personales, académicos, económicos, sociales e institucionales; los factores con mayor frecuencia de uso se encuentran en la dimensión personales, donde los más investigados la edad, etnia, género. En la dimensión académica, el factor más referenciado fue la prueba de ingreso a la universidad, considerado en varias investigaciones como un predictor importante en la deserción escolar.

Las técnicas de minería de datos utilizadas para resolver el problema de la deserción en las universidades se clasificaron en las siguientes: estadísticas, inteligencia artificial y combinada. De ellas, Decision Tree classifier y Support Vector Machine fueron las más utilizadas y presentan una tasa de precisión del 98 % y 97.5 %, respectivamente. Finalmente, existen varias herramientas de minería de datos utilizadas para la predicción, pero la más usada es WEKA, debido a su flexibilidad para el modelamiento predictivo.

CAPÍTULO IV: FACTORES DE DESERCIÓN DESDE LA PERSPECTIVA DEL ESTUDIANTE

4.1 Introducción

El abandono universitario es uno de los problemas que afecta a la mayoría de las instituciones de educación superior de todo el mundo, así, convierte en un tema que genera controversia en el entorno educativo en el que se involucran los administradores, profesores y estudiantes (Díaz, 2008). En la actualidad, las altas tasas de deserción son consideradas como posibles deficiencias en el sistema educativo de pregrado (Castaño, Gallón, Gómez & Vásquez, 2008). Esto se puede evidenciar en los indicadores académicos y de gestión administrativa que presentan los administradores de las universidades a las entidades de control de la educación superior. En el año 2016, en la India la tasa de deserción fue del 15.9 % (Himmel, 2002), mientras que en Bélgica corresponde al 26,9 %, en Estados Unidos la tasa de abandono de primero y segundo año alcanzó el 44.8 % (Willging & Johnson, 2009) y en Costa Rica superó el 49 %. Según la Organización de las Naciones Unidas (ONU) en el año 2016, en algunos países latinoamericanos como en Colombia y Ecuador, la deserción superó el 40 % y en Brasil llegó aproximadamente al 54 %.

Se puede considerar que la deserción está influenciada por un conjunto de factores que interactúan entre sí y que ejercen influencia negativa sobre la decisión que toman los estudiantes en la culminación de sus estudios universitarios (Hovdhaugen, 2009). La revisión de la literatura ha permitido identificar trabajos relacionados con factores que influyen en la deserción, sin embargo, en general, estas no consideran la perspectiva del estudiante, quien es el actor principal de la deserción. Por lo que este trabajo basado en las teorías organizacionales y de educación propone 11 nuevos factores que afectan a la deserción estudiantil universitaria desde la perspectiva del

estudiante, para efectos de comprobar su influencia, se realiza un estudio empírico sobre 3773 estudiantes.

4.2 Trabajos relacionados

La deserción universitaria es considerada como la suspensión de un proceso académico y la no culminación de un proceso formativo en la Educación Superior (Gartner, Dussán & Montoya, 2016). Es un problema que causa preocupación entre los gobiernos y se ha convertido en una debilidad en el nivel de educación universitario debido a sus altas tasas, que tiene efectos negativos en los estudiantes y en el crecimiento económico de la sociedad (Gitto, Minervini & Monaco, 2016). En este mismo sentido, se considera como desertor a un estudiante que no presenta actividad académica durante dos semestres consecutivos (Gartner, Dussán & Montoya, 2016). Para Stratton, *et al.* (2008), la mayor parte de los estudios relacionados a la deserción indican que una vez que el alumno ha desertado, este jamás regresa a la universidad, por lo tanto, la deserción se considera como permanente.

La deserción de los estudiantes es de naturaleza multifactorial (O'Neill, Christensen, Vonsild & Wallstedt, 2014), abarca dimensiones académicas personales, sociales, institucionales y económicas, como las propuestas por Sánchez (2002). Los factores tienen un impacto significativo en la probabilidad de que se produzca la deserción (Gitto, Minervini & Monaco, 2016) e imponen en muchas ocasiones restricciones externas en los cuales se comprometen las capacidades de los estudiantes para completar sus estudios. (Human-Vogel & Rabe, 2015)

La revisión de la literatura presenta evidencia empírica y teórica que busca identificar factores que inciden en la deserción. En las tablas 18 a la 22 se presenta 112 factores que influyen en la deserción estudiantil universitaria, los cuales se encuentran clasificados en cinco dimensiones.

Tabla 18
Dimensión institucional

Factor	Referencia
Ambiente del campus	Sivakumar, Venkataraman & Selvaraj (2016)
Tipo de escuela secundaria	Guarín, Guzmán & González (2015)
Involucramiento institucional	Belo & Oliveira (2015)
Infraestructura de la universidad	Sivakumar, Venkataraman & Selvaraj (2016)

Tabla 19
Dimensión personal

Factor	Referencia
Ajuste	Sivakumar, Venkataraman & Selvaraj (2016)
Edad	Yukselturk, Ozekes & Turel (2014), Wray, Brett, Aspland & Gardiner (2012), Tan & Shao (2015), Levy (2007), Oeda & Hashimoto (2017), Park & Choi (2009), Houdhaugen (2009), Chen & Desjardins (2008), Herzog (2005), Tumen, Shulruf & Hattie (2008), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015), Willging & Johnson (2009), Arulampalam, Naylor & Smitg (2007), González, Heracleous & Winters (2012), Stratton, O'Toole & Wetzel (2008), Arulampalam, Naylor & Smith (2005), Di Pietro & Cutillo (2008), Melguizo, Sanchez & Velasco (2016)
Cambio de meta	Saranya & Rajeswari (2016), Sivakumar, Venkataraman & Selvaraj (2016)
Elección para cambiar el curso actual	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
País o ciudad de origen	Hoffait & Schyns (2017), Guarín, Guzmán & González (2015), Heredia, Amaya & Barrientos (2015)
Dependientes	Wray, Brett, Aspland & Gardiner (2012)
Discapacidad	Wray, Brett, Aspland & Gardiner (2012)
Dirección	Wray, Brett, Aspland & Gardiner (2012), Lam-ON & Boongoen (2014), Levy (2007), Di Pietro & Cutillo (2008), Sivakumar, Venkataraman & Selvaraj (2016), Tan & Shao (2015)
Apoyo de los padres	Arulampalam, Naylor & Smitg (2007)
Compromiso del estudiante	Houdhaugen (2009), Reschly & Christenson (2006), Willcoxson, Cotter & Joy (2011)

Tabla 19
Dimensión personal (Continuación)

Factor	Referencia
Etnia	Wray, Brett, Aspland & Gardiner (2012), Thammasiri, Delen, Meesad & Kasap (2014), Tan & Shao (2015), Chen & Desjardins (2008), Herzog (2005), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Guarín, Guzmán & González (2015), Martinho, Nunes & Minussi (2013), Willging & Johnson (2009), Arulampalam, Naylor & Smitg (2007), González, Heracleous & Winters (2012), Oseguera & Rhee (2009)
Género	Yukselturk, Ozekes & Turel (2014), Wray, Brett, Aspland & Gardiner (2012), Yasmin (2013), Tan & Shao (2015), Yi. et al. (2015), Natek & Zwilling (2014), Lam-ON & Boongoen (2014), Paura & Arhipova (2014), Hoffait & Schyns (2017), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Levy (2007), Jadric, Garaca & Cukusic (2010), Park & Choi (2009), Houdhaugen (2009), Sultana, Khan & Abbas (2017), Chen & Desjardins (2008), Herzog (2005), Tumen, Shulruf & Hattie (2008), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Iepsen, Bercht & Reategui (2013), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015), Willging & Johnson (2009), Melguizo, Sanchez & Velasco (2016)
Tiene un computador	Martinho, Nunes & Minussi (2013)
Problemas de salud	Saranya & Rajeswari (2016)
Nivel de interés	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Motivación intrínseca	Arifin (2016), Willcoxson, Cotter & Joy (2011)
Liderazgo	Sultana, Khan & Abbas (2017)
Nivel de compromiso	Human_Vogel & Rabe (2015)
Vive en el campus	Herzog (2005)
Soledad	Alkan (2014)
Estado civil	Jia & Maloney (2015), Yasmin (2013), Jadric, Garaca & Cukusic (2010), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015), Stratton, O'Toole & Wetzel (2008)
Medida de persistencia del alumno	Duarte, Ramos, Goncalves (2014)
Pesimismo	Oeda & Hashimoto (2017), Iepsen, Bercht & Reategui (2013)
Lugar de residencia	Tan & Shao (2015), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Herzog (2005)
Autoeficacia	Hoffait & Schyns (2017), Willcoxson, Cotter & Joy (2011), Arifin (2016), Yukselturk, Ozekes & Turel (2014)
Satisfacción del estudiante	Saranya & Rajeswari (2016), Levy (2007), Duque (2014), Alkan (2014)
Fuente de matrícula	Tan & Shao (2015)
Implicación vocacional	Belo & Oliveira (2015)
Experiencia	Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)

Tabla 20
Dimensión económica

Factor	Referencia
Beca	Lin (2012),Thammasiri, Delen, Meesad & Kasap (2014), Hoffait & Schyns (2017)
Pertenece al cuartil de pobreza	Yasmin (2013)
Trabaja en el campus	Herzog (2005)
Dependencia	Arulampalam, Naylor & Smitg (2007)
Préstamo estudiantil	Thammasiri, Delen, Meesad & Kasap (2014)
Ingresos familiares	Martinho, Nunes & Minussi (2013)
Trabajo del padre	Jia & Maloney (2015)
Preocupaciones financieras	Oseguera & Rhee (2009),Willcoxson, Cotter & Joy (2011), Lin (2012)
Inversión	Human-Vogel & Rabe (2015)
Ingreso bruto de los padres	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Préstamos recibidos	Lin (2012), Thammasiri, Delen, Meesad & Kasap (2014), Chen (2012)
Estado de empleo del estudiante	Sultana, Khan & Abbas (2017)
Tasa de estudiantes	Arulampalam, Naylor & Smith (2005)
Tipo de ayuda financiera	Thammasiri, Delen, Meesad & Kasap (2014), Paura & Arhipova (2014), Chen (2012), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)

Tabla 21
Dimensión académica

Factor	Referencia
Absentismo	Sultana, Khan & Abbas (2017)
Habilidad académica	Di Pietro & Cutillo (2008),Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012)
Sobrecarga académica	Di Pietro & Cutillo (2008)
Desempeño académico	Jia & Maloney (2015), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Melguizo, Sanchez & Velasco (2016)
Edad de ingreso	Al-Barrak & Al-Razgan (2015)
Resultado de la evaluación formativa	Tan & Shao (2015)
Mejor desempeño en el examen GPA, SAT	Lin (2012), Jia & Maloney (2015), Thammasiri, delen, Meesad & Kasap (2014), Duarte, Ramos, Goncalves (2014), Aulck, Velagapudi, Blumenstock & West (2016), Lam-ON & Boongoen (2014), Levy (2007), Huang & Fang (2013), Koonce & Hening (2009), Chen (2012), Reschly & Christenson (2006), Alkan (2014),Chen & Desjardins (2008), Herzog (2005), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Guarín, Guzmán & González (2015), Willging & Johnson (2009), Oseguera & Rhee (2009), Human-Vogel & Rabe (2015)
Cohorte	Wray, Brrett, Aspland & Gardiner (2012), Yi. et al. (2015), Willging & Johnson (2009), González, Heracleous & Winters (2012)

Tabla 21
 Dimensión académica (Continuación)

Factor	Referencia
Participación curricular	Martinho, Nunes & Minussi (2013)
Título	Jia & Maloney (2015), Wray, Brrett, Aspland & Gardiner (2012), Natek & Zwilling (2014), Chen (2012), Jadric, Garaca & Cukusic (2010), Chen & Desjardins (2008)
Aspiración de grado	Oseguera & Rhee (2009)
Duración del programa de grado	Arulampalam, Naylor & Smith (2005)
Intensión de desertar	Alkan (2014)
Meta educativa	Houdhaugen (2009)
Idioma Inglés	Lykountzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)
Inscrito en otra institución	Saranya & Rajeswari (2016)
Calificaciones de ingreso	Wray, Brrett, Aspland & Gardiner (2012), Duarte, Ramos, Goncalves (2014)
Experiencia	Yukselturk, Ozekes & Turel (2014), Wray, Brrett, Aspland & Gardiner (2012)
Examen final	Tan & Shao (2015), Cukusic, Garaca & Jadric (2014), Xing, Chen, Stein & Marcinkowski (2016), Huang & Fang (2013), Sultana, Khan & Abbas (2017), Al-Barrak & Al-Razgan (2015)
Examen parcial	Al-Barrak & Al-Razgan (2015)
Créditos del primer semestre	Herzog (2005)
Motivo de elección de la carrera	Houdhaugen (2009)
Cuestionario	Al-Barrak & Al-Razgan (2015)
Participa en actividades extracurriculares	Saranya & Rajeswari (2016), Sivakumar, Venkataraman & Selvaraj (2016)
Calificaciones de la secundaria	Jadric, Garaca & Cukusic (2010), Sultana, Khan & Abbas (2017), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Progreso en los estudios	Wray, Brrett, Aspland & Gardiner (2012)
Preparación	Yukselturk, Ozekes & Turel (2014)
Créditos reconocidos	Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Tan & Shao (2015), Heredia, Amaya & Barrientos (2015)
Uso de recursos	Belo & Oliveira (2015)
Satisfacción del curso	Sivakumar, Venkataraman & Selvaraj (2016)
Puntuación de integración académica	Chen (2012), Chen & Desjardins (2008), Al-Barrak & Al-Razgan (2015)
Calificaciones	Paura & Arhipova (2014), Hoffait & Schyns (2017), Heredia, Amaya & Barrientos (2015), Arulampalam, Naylor & Smitg (2007), Al-Barrak & Al-Razgan (2015), Melguizo, Sanchez & Velasco (2016)
Autoevaluación	Sultana, Khan & Abbas (2017), Martinho, Nunes & Minussi (2013)
Estado de matrícula de estudiantes	Sultana, Khan & Abbas (2017), Arulampalam, Naylor & Smith (2005), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)

Tabla 21
 Dimensión académica (Continuación)

Centro de estudios	Tan & Shao (2015), Martinho, Nunes & Minussi (2013), Arulampalam, Naylor & Smitg (2007), Stratton, O'Toole & Wetzel (2008)
Nivel de estudios	Tan & Shao (2015), Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009), Park & Choi (2009), Herzog (2005)
Jornada de estudios	Martinho, Nunes & Minussi (2013)
Tasa de éxito	Jia & Maloney (2015), Tan & Shao (2015)
Apoyo para el aprendizaje	Willcoxson, Cotter & Joy (2011)
Total de cursos reprobados	Tan & Shao (2015)

Tabla 22
 Dimensión social

Factor	Referencia
Alojamiento en el campus	Arulampalam, Naylor & Smitg (2007), Arulampalam, Naylor & Smith (2005), Oseguera & Rhee (2009)
Categoría marginada o sección vulnerable de la sociedad	Yasmin (2013), González, Heracleous & Winters (2012)
Estado de la Universidad	Levy (2007)
Soprote comunitario	Sultana, Khan & Abbas (2017)
Estado del empleo	Yasmin (2013), Natek & Zwillling (2014), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Martinho, Nunes & Minussi (2013)
Problemas familiares	Sivakumar, Venkataraman & Selvaraj (2016)
Composición familiar	Sivakumar, Venkataraman & Selvaraj (2016)
Nivel de educación del padre	Jia & Maloney (2015), Jadric, Garaca & Cukusic (2010), Houdhaugen (2009), Chen & Desjardins (2008), Martinho, Nunes & Minussi (2013), Heredia, Amaya & Barrientos (2015)
Indicador de vivienda	Jadric, Garaca & Cukusic (2010)
Nivel de implicación de las redes sociales	Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015)
Medio de transporte	Martinho, Nunes & Minussi (2013)
Migrante	Yi, et al. (2015), Houdhaugen (2009)
Nivel de educación de la madre	Jia & Maloney (2015), Jadric, Garaca & Cukusic (2010), Houdhaugen (2009), Sultana, Khan & Abbas (2017), Chen & Desjardins (2008), Martinho, Nunes & Minussi (2013), Melguizo, Sanchez & Velasco (2016)
Ocupación de la madre, padre	Heredia, Amaya & Barrientos (2015), Willging & Johnson (2009)
Ocupación de los parientes	Saranya & Rajeswari (2016)
Estatus político	Tan & Shao (2015)
Estatus social	Jadric, Garaca & Cukusic (2010), Park & Choi (2009), Heredia, Amaya & Barrientos (2015), Arulampalam, Naylor & Smitg (2007), Belo & Oliveira (2015), Arulampalam, Naylor & Smith (2005)
Stress	Saranya & Rajeswari (2016)

4.3 Metodología

4.3.1 Nuevos factores identificados

El conjunto de nuevos factores se construyó sobre el análisis de la revisión de la literatura, 65 teorías organizacionales, 12 teorías educativas y el raciocinio lógico. Como resultado del estudio se obtuvo 11 factores, los cuales se describen a continuación.

F1: Conocimiento limitado sobre el uso del software especializado en la carrera. Se refiere a la capacidad que tiene el estudiante para utilizar software especializado con fines académicos. Este tipo de software está diseñado con la finalidad de mejorar los conocimientos de los estudiantes mediante la aplicación de tecnologías y no se relaciona con material didáctico multimedia, entornos virtuales, internet, uso de blogs, wikis, foros, chats, mensajería, entre otros. Es importante analizar su influencia en la deserción universitaria, debido a que un estudiante debe poseer el conocimiento adecuado en el uso de las tecnologías aplicadas en su proceso de formación universitaria. El uso del software aplicado al proceso de enseñanza facilita el entendimiento y en muchas ocasiones se vuelve necesario para el dominio de un área de estudio específica. Por ejemplo, en la Carrera de Ciencias Computacionales, el uso de la tecnología como PHP, Java, SQL, MySQL se vuelve necesaria para el desarrollo de sistemas informáticos. Si un estudiante no domina la tecnología como parte de su proceso de formación académica se siente desmotivado y deja de atender sus actividades académicas, por lo que pierde el control de los resultados y dominios del aprendizaje.

Por otro lado, cuando el alumno se siente motivado, muestra interés en su aprendizaje y en las actividades que conducen a su desarrollo académico. Teoría de la motivación. (Martín, Martínez, Nieto & Núñez, 2017)

Hipótesis H1: El conocimiento limitado sobre el uso del software especializado en la carrera influye en la deserción estudiantil universitaria.

F2: Embarazo planificado no planificado. Se refiere al proceso de gestación de una estudiante en el transcurso de su formación académica universitaria. En la adolescencia es considerado como un problema biomédico con elevado riesgo de complicaciones (Magadi, Otineo, Obare & Taffa, 2004), tiene implicancias socioculturales y psicológicas que afectan al estudiante en lo personal, educacional, familiar y en el ámbito social (Mueller, *et al.*, 2017). Por consiguiente, la estudiante embarazada limita sus oportunidades de estudio y trabajo, puesto que la mayoría de ellas se ve obligada a desertar del sistema académico, lo que perjudica sus condiciones de vida. La paternidad en la juventud, especialmente en el contexto secundario, genera muchas desventajas para la consecución de los objetivos académicos (Black, Fleming & Rome, 2012), este problema se enfatiza más aun en la universidad puesto que el nivel de exigencia académico es mayor y, por lo tanto, requiere de más atención en el desarrollo de las actividades.

Hipótesis H2: *El embarazo planificado/no planificado influye en la deserción estudiantil universitaria.*

F3: Compromiso del docente con el estudiante. Se refiere a la afirmación de principios que el docente expresa en el desarrollo de la formación académica y humana de sus estudiantes. La comprensión de la competencia docente expresa las potencialidades para orientar su trabajo con iniciativa, flexibilidad y autonomía (Hallinger, Hosseingholizadeh, Hashemi & Kouhsari, 2018). Es relevante en la deserción estudiantil universitaria, ya que a partir de la integración de conocimientos, habilidades, motivación y valores que se expresan en un desempeño docente eficiente, ético y con compromiso social, el estudiante universitario se sentirá motivado de continuar con sus estudios. Los programas de tutorías y mentorías insertos en las universidades afirman la importancia que tiene este factor debido a que el acompañamiento académico y estudiantil se aplica para reducir las tasas de deserción. Por otro lado, un profesional competente que posee conocimientos y habilidades que le posibilitan desempeñarse con éxito en una profesión específica (teoría Gestalt, (Miranda, 2017)), permitirá el desarrollo humano y académico de los

alumnos, así los motivará a alcanzar con éxito la culminación de sus estudios universitarios.

***Hipótesis H3:** El factor compromiso del docente con el estudiante influye en la deserción estudiantil universitaria.*

F4: Compromiso económico del hijo primogénito con la familia. Se refiere a la obligación que adquiere el mayor de los hijos varones de una familia y que, en ausencia del padre, es quien asume las responsabilidades económicas con sus familiares (Ito, 2009). Es relevante en la deserción estudiantil universitaria, debido a que el rango de orden de nacimiento entre hijos e hijas afecta las competencias en el desarrollo académico de los estudiantes (Yu & Su, 2006), especialmente cuando los hijos mayores tienen que abandonar sus estudios para trabajar y poder sustentar las necesidades económicas de sus hermanos. Cabe destacar que las personas se sienten motivadas cuando cumplen sus necesidades prioritarias antes de satisfacer las necesidades de crecimiento profesional (teoría de Maslow, (Leal & Torrealba, 2017)), es por ello que, si un estudiante no atiende las necesidades básicas de su familia como vivienda, salud, alimentos, entre otros, se sentirá desmotivado, lo que influirá en la decisión de abandonar la universidad.

***Hipótesis H4:** El compromiso económico con la familia por ser hijo primogénito influye en la deserción estudiantil universitaria.*

F5: Bullying. Se refiere al patrón de comportamiento agresivo que se da entre pares e implica acciones no deseadas, negativas y repetidas en el tiempo (Rojas, 2013), el abuso entre iguales es una conducta de persecución física o psicológica (Dueñas & Valdivia, 2017), por consiguiente, si un estudiante se siente afectado física y emocionalmente, su proceso de aprendizaje también se verá afectado. Este problema es muy estudiado en la educación secundaria, sin embargo, en la educación universitaria causa mayores inconvenientes debido a que en su mayoría los estudiantes viven solos, fuera del cuidado y protección de sus padres, lo que imposibilita tener ayuda temprana que les

permita resolver los problemas de agresión. El bloqueo social en el que encuentra inmerso y su necesidad de dejar el ambiente hostil pueden conducir al estudiante a abandonar las aulas universitarias. Así mismo, la continuidad de estas relaciones distorsionadas provoca en las víctimas efectos claramente negativos como el descenso de su autoestima, estados de ansiedad e incluso cuadros depresivos (teoría de la personalidad, (Di & Rayan, 2017)).

Hipótesis H5: *El factor bullying influye en la deserción estudiantil universitaria.*

F6: Machismo/feminismo. Se refiere a aquella actitud o manera de pensar de quien sostiene que el varón es por naturaleza superior a la mujer, o que la mujer es por naturaleza superior al hombre (Garrido, 2016). Es relevante en la deserción universitaria debido a que el comportamiento de los estudiantes puede presentar diferencias en cuanto a los roles de género, especialmente en las carreras en las que existe mayor cantidad de estudiantes mujeres y limitada cantidad de estudiantes varones o viceversa. Por consiguiente, la relación negativa con el entorno social en el que se desarrollan los individuos puede provocar la degradación de la imagen del hombre y la mujer en el contexto social y académico (teoría de la socialización).

Hipótesis H6: *El machismo/feminismo influye en la deserción estudiantil universitaria.*

F7: Vicios adquiridos por el estudiante. Se refiere a las conductas negativas de los estudiantes que son socialmente aceptadas y tienen la característica común de proveer un estado de gratificación inmediata, por ejemplo, adicción a las redes sociales, drogas, juegos, entre otras. Esto puede deberse a que los estudiantes cuando están en su etapa de formación profesional se ven limitados a establecer diariamente relaciones con amigos o familiares que hace que se aíslen y caigan en situación de tedio y se vuelven adictos a algo.

Los seres humanos sienten una profunda necesidad de establecer relaciones y conexión (teoría del modelamiento mental e inducción, (Jantke & Drefahl, 2016)). Cuando la relación entre una adicción y el estudiante es constante comienza a ser problemática, debido a que el estudiante prioriza la actividad adictiva en vez de sus obligaciones académicas.

Hipótesis H7: *Vicios que adquiere el estudiante influye en la deserción estudiantil universitaria.*

F8: Hijos del estudiante. Se refiere a la cantidad de hijos que el estudiante tiene en el transcurso de su formación universitaria. La fecundidad y la reproducción entre los estudiantes de secundaria es un problema constante que mantiene relación con la permanencia y deserción del sistema educativo (Rodríguez, 2005), esto se debe a que los estudiantes tienen que dedicar tiempo para el cuidado de los hijos y a su vez deben conseguir el sustento económico que les permita mantener las condiciones necesarias para el bienestar de los hijos. Este problema se agudiza en el contexto de educación universitario debido a que el alumno tiene horarios de estudios más rígidos que requiere mayor responsabilidad y compromisos de tiempo para desarrollar sus actividades académicas.

Hipótesis H8: *El número de hijos del estudiante influye en la deserción estudiantil universitaria.*

F9: Adaptación del estudiante a las metodologías de formación universitaria. Se refiere al grado de ajuste que tiene el estudiante universitario frente a un cambio de los métodos de aprendizaje que aplican los docentes y que en el contexto universitario difieren de los aplicados en la secundaria. Es relevante en la deserción universitaria debido a que el alumno debe ser capaz de manipular el conocimiento, ponerse al día en los contenidos académicos, seleccionar los espacios apropiados que permitan desarrollarse profesionalmente y adaptarse a los cambios.

Hipótesis H9: *Adaptación del estudiante a las metodologías de formación universitaria influye en la deserción estudiantil universitaria.*

F10: Ranking de una carrera o institución. Se refiere a la posición que adquiere una universidad o carrera respecto a otra. El ranking universitario se presenta en listas clasificadas por grupos de instituciones que son evaluadas de acuerdo criterios e indicadores de calidad (Yoguez, 2009). Este tipo de clasificaciones pueden tener impacto en el prestigio de la universidad e influencia en el número y en la calidad de los estudiantes (Giannoulis, 2010). Se puede señalar que la percepción de prestigio que tiene el estudiante sobre la universidad aumentará el interés de permanecer en ella, por el incremento de oportunidades de trabajo, estatus social, beneficios académicos y de bienestar estudiantil. Sin embargo, si durante el transcurso de los estudios se va disolviendo el interés del estudiante disminuye, en algunos casos podría abandonar sus estudios para buscar mejores beneficios en otras universidades.

Hipótesis H10: *El ranking de una carrera o institución influye en la deserción estudiantil universitaria.*

F11: Percepción del estudiante sobre la inserción en el campo laboral. Se refiere a la idea o proyección que tiene el estudiante sobre la formación y recursos académicos que brinda la universidad para insertarse en el ámbito laboral. Existe estrecha relación entre la educación superior y el mundo laboral, debido a que esta proporciona al estudiante el conocimiento, las habilidades y competencias que aumentan la capacidad del alumno para conseguir un empleo (García-Blanco & Sempértegui, 2018). El vínculo educación-trabajo genera impacto positivo o negativo en el estudiante. Se considera que a mayor nivel educativo mayores posibilidades de éxito en el campo laboral, que traerán consigo mejor estatus económico e influencia social. (Navarro, 2014)

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

Tabla 23
Nuevos factores de deserción universitaria identificados

ID	Factor	Descripción	Sustento
F01	Conocimiento limitado sobre el uso de software especializado de la carrera	Capacidad que tiene el estudiante para utilizar software especializado con fines académicos	Teoría de la motivación (Martín, Martínez, Nieto & Núñez, 2017)
F02	Embarazo planificado/no planificado	Proceso de gestación de una estudiante en el transcurso de su formación académica universitaria	(Taffa & Obare, 2004), (Black, Fleming & Rome, 2012)
F03	Compromiso del docente con el estudiante	Afirmación de principios que el docente expresa en el desarrollo de la formación académica y humana de sus estudiantes	Teoría de Gestalt (Miranda, 2017)
F04	Compromiso económico del hijo primogénito con la familia	Obligación que adquiere el mayor de los hijos varones de una familia, y que, en ausencia del padre, es quien asume las responsabilidades económicas con sus familiares (Ito & Caste, 2009)	Teoría Maslow (Leal & Torrealba, 2017)
F05	Bullying	Agresión física, verbal y psicológica que nace de la intención de dañar a otra persona (Rojas, 2013)	Teoría de la personalidad (Di & Rayan, 2017)
F06	Machismo/feminismo	Actitud o manera de pensar de quien sostiene que el varón es por naturaleza superior a la mujer o que la mujer es por naturaleza superior al hombre (Garrido, 2016)	(Garrido, 2016)
F07	Vicios adquiridos por el estudiante	Conductas negativas de los estudiantes que son socialmente aceptadas y tienen la característica común de proveer un estado de gratificación inmediata, por ejemplo, adicción a las redes sociales, drogas, juegos, entre otras	(Jantke & Drefahl, 2016)
F08	Número de hijos del estudiante	Cantidad de hijos que el estudiante tiene en el transcurso de su formación universitaria	(Rodríguez, 2005)
F09	Adaptación de estudiante a las metodologías de formación universitaria	Grado de ajuste que tiene el estudiante universitario frente a un cambio de los métodos de aprendizaje que aplican los sus docentes y que en el contexto universitario difieren de los aplicados en la secundaria	
F10	Ranking de la Universidad o Carrera	Posición que adquiere una universidad o carrera respecto a otra (Yoguez, 2009)	(Giannoulis, 2010)
F11	Perspectivas del estudiante sobre la inserción en el campo laboral	Idea o proyección que tiene el estudiante sobre la formación y recursos académicos que brinda la universidad para insertarse en el ámbito laboral	(García-Blanco & Sempértegui, 2018), (Navarro, 2014)

El modelo para determinar nuevos factores de deserción incluye un conjunto de once factores que pueden ser las causas en el abandono universitario (Ver figura 6).

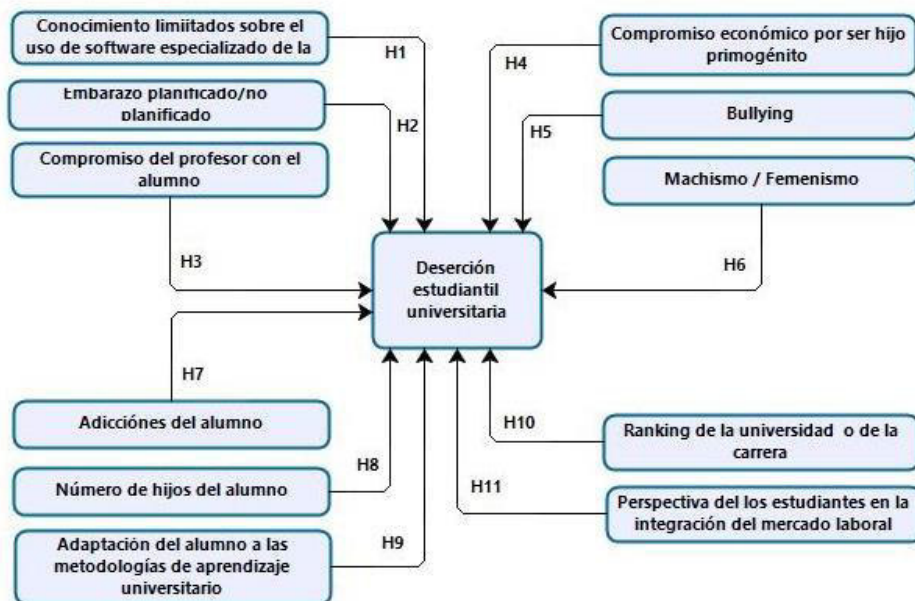


Figura 6. Modelo Conceptual

Fuente: Elaborado por el autor

4.3.2 Recopilación de Información

Se aplica una encuesta basada en el modelo conceptual propuesto mediante la herramienta Google Form por más de tres meses, desde junio hasta octubre de 2017. Estuvo dirigida a los estudiantes matriculados en el nivel de pregrado de Universidades Públicas del Ecuador desde marzo del 2012 hasta octubre de 2017, de los cuales 3773 alumnos respondieron la encuesta. El objetivo de la aplicación del instrumento fue determinar el punto de vista de los estudiantes universitarios sobre las posibles causas que inciden en la deserción universitaria.

La encuesta estuvo estructurada en 3 secciones. Sección 1: información relacionada con la caracterización de la institución y de los estudiantes (10 preguntas), sección 2: preguntas que determinaron los factores que inciden en

la deserción (13 preguntas), sección 3: preguntas que complementaron el estudio (7 preguntas). Las preguntas de la sección 2 fueron evaluadas con la escala de Likert de 5 puntos que corresponden a 1: no influye, 2: baja influencia, 3: mediana influencia, 4: alta influencia, 5: influye totalmente (ver anexo 1).

Para la validación de las preguntas de la encuesta, se estableció dos pruebas piloto, la primera estuvo dirigida a personas con experiencia en temas relacionados con la deserción en las universidades y se aplicó a una muestra de 200 estudiantes, el objetivo fue verificar la relación existente de las preguntas con las hipótesis planteadas. Como resultado se corrigió la redacción y se plantean las preguntas bajo un lenguaje más sencillo que pueda ser entendido por los estudiantes.

4.4 Resultados

Se estructura en tres etapas:

1. Análisis descriptivo de la población, utilizado para conocer las características demográficas de los estudiantes.
2. Prueba de fiabilidad y validez del cuestionario, mediante Alpha de Cronbach para determinar la consistencia interna del instrumento y asegurar que los ítems medidos en el instrumento mediante la escala de Likert estén altamente correlacionados.
3. Modelamiento de datos, mediante regresión lineal, basado en la metodología de Box Jenkins.

4.4.1 Análisis descriptivo de la población

La tabla 24, presenta un resumen de las características demográficas de los estudiantes matriculados y que fueron utilizados como sujetos de prueba.

Tabla 24
Estadística descriptiva

Variable	Descripción	Cantidad	Porcentaje
Género	Hombre	1745	46.25%
	Mujer	2028	53.75%
Grupos de edad	Age < 20	2171	57.54%
	Age 21-28	1494	39.60%
	Age > 29	108	2.86%
	Casado	521	13.81%
Estado civil	Divorciado	33	0.87%
	Soltero	3032	80.36%
	Unión Libre	187	4.96%
	Afro descendiente	6	0.16%
Etnia	Blanco	19	0.50%
	Indígena	142	3.76%
	Mestizo	3581	94.91%
	Montubio	11	0.29%
Tiene empleo	Mulato	13	0.34%
	No	2763	73.23%
	Si	1010	26.77%
	Madre y hermanos	599	15.88%
Tipo de Hogar	Padre y hermanos	750	19.88%
	Padre, madre y hermanos	1926	51.05%
	Esposa e hijos	303	8.03%
	Otros familiares	195	5.17%
	Ninguna	362	9.59%
	Primaria	1816	48.13%
Nivel de educación del padre	Secundaria	1147	30.40%
	Superior	448	11.87%
	Ninguna	336	8.91%
Nivel de educación de la madre	Primaria	1889	50.07%
	Secundaria	1177	31.20%
	Superior	371	9.83%

Los resultados de la indagación socioeconómica de los estudiantes permitieron determinar una serie de diferencias significativa en la muestra utilizada para el estudio, por ejemplo, un porcentaje mayor de la población femenina, estudiantes que provienen de hogares incompletos sin padre o madre y que viven al cuidado de otros familiares, nivel de formación del padre y madre en el que predomina el nivel primario, y porcentajes altos de estudiantes que mantienen obligaciones laborales.

3.4.2 Fiabilidad de los datos

Se utilizó el estadístico Alfa de Cronbach, y mediante el método de varianza de los ítems se determinó el índice de consistencia interna del instrumento.

Los resultados obtenidos a través del software SPSS arrojan un valor de 0.908, lo cual indica que el cuestionario aplicado es fiable.

Tabla 25
Resultado de la validez de la encuesta

Estadísticos de fiabilidad	
Alfa de Cronbach	N de elementos
.908	37

Por otro lado, la tabla 26 presenta el resultado de la valoración que dieron los estudiantes en la escala de Likert sobre los factores que influyen en la deserción estudiantil universitaria.

Para el análisis de las respuestas del instrumento aplicado, se estableció la media de las puntuaciones del total de las respuestas. Mediante el método de agrupación visual, se obtuvieron los puntos de corte de los intervalos entre 1 y 5 puntos, esto permitió establecer los límites superiores de los rangos.

Tabla 26
Resultado de la valoración de la Escala de Likert

Escala	Frecuencia	%	% Válido	% Acumulado
Baja influencia	14	0.4	0.4	0.4
Mediana influencia	110	3.5	3.5	3.9
Alta influencia	1116	25.5	25.5	29.4
Influye totalmente	2533	70.6	70.6	100.0
Total	3773	100.0	100.0	

Por otro lado, la tabla 27 presenta el cálculo de las medidas de tendencia central de dispersión como la media, moda y desviación estándar respecto a las respuestas de los estudiantes sobre la percepción de la deserción en las instituciones de educación superior, el promedio de la media >4 indica que existe una alta relación entre las respuestas de los estudiantes respecto a la deserción.

Tabla 27

Medidas de tendencia central sobre la percepción de la deserción

Hypothesis	Mean	Moda	Desv.tip
H1	4.27	5	0.889
H2	4.28	5	1.004
H3	4.17	5	1.085
H4	4.35	5	1.042
H5	4.02	4	0.914
H6	4.03	4	1.022
H7	4.05	4	0.892
H8	4.25	5	1.078
H9	4.24	5	1.173
H10	4.21	5	1.138
H11	4.33	5	0.991

4.4.3 Modelado de los datos

Para la construcción del modelo, se utilizó la metodología de Box Jenkins propuesta por Gujarati & Porter (2010) que consta de 4 etapas: identificación, estimación, verificación y pronóstico, descritas a continuación.

Especificación: El modelo cuenta con un vector de variables independientes $X = (F_1, F_2, \dots, F_n)$ que son los factores, las cuales se consideran como influencia en el resultado de Y (deserción), se parte de la fórmula (1) propuesta por Del Carmen (2010).

$$P(y = 1 | x) = \frac{e^{B_0 + B_1x}}{1 + e^{B_0 + B_1x}} \quad (1)$$

Donde:

B_0 = término independiente

P = probabilidad de ocurrencia del evento

Y = variable dependiente

X = variables independientes (F)

B_1x = combinación de variables independientes

Estimación: En esta etapa se estima los coeficientes de significancia y correlación de factores, los cuales se presentan en la Tabla 10. Se utilizó regresión logística por ser un método de estimación no lineal de parámetros, y porque permite predecir la probabilidad de ocurrencia de un evento dadas

las características de las variables independientes [83]. Una de las características del modelo logit es que la variable dependiente debe contener datos binarios, esto significa que puede tener dos posibles resultados:

Y=1 Alta probabilidad de influencia de deserción
Y=0 Baja probabilidad de influencia de deserción

La prueba estadística para determinar la influencia de los factores identificados en la deserción estudiantil universitaria se la realizó a través del software Eviews. Los resultados del modelo Logistic Regression se presentan en la ecuación (2).

$$Y = 1 - @CLOGISTIC(-(0.818043 * F1 + 0.629028 * F2 - 0.390826 * F3 + 0.380309 * F4 - 0.809717 * F5 + 0.852620 * F6 + 0.687710 * F7 + 1.384878 * F8 + 0.281206 * F9 + 0.637297 * F10 + 1.346561 * F11 - 4.819381)) \quad (2)$$

La tabla 28, permite identificar el comportamiento de cada factor y su relación positiva (+) o negativa (-) respecto a la deserción.

Tabla 28
Coefficiente de significancia de los factores

Dependent Variable: DES				
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)				
Included observations: 3773				
Convergence achieved after 3 iterations				
Coefficient covariance computed using observed Hessian				
Variable	Coefficient	Std. Error	z-Statistic	P-value
C	-4.819.381	0.356954	1.350.141	0.00000
F1	0.818043	0.187763	4.356.790	0.00000
F2	0.629028	0.180224	3.490.262	0.00050
F3	-0.390826	0.139217	2.807.320	0.00400
F4	0.380309	0.172667	2.202.562	0.02760
F5	0.809717	0.181720	4.455.839	0.00000
F6	0.852620	0.152875	5.577.225	0.00000
F7	0.687710	0.150909	4.557.112	0.00000

Tabla 28
Coefficiente de significancia de los factores (Continuación)

F8	1.384.878	0.160912	8.606.428	0.00000
F9	0.281206	0.135480	2.075.625	0.03790
F10	0.637297	0.143733	4.433.882	0.00000
F11	1.346.561	0.202286	6.656.711	0.00000
Prob(LR statistic)	0.00000			

El valor de $P(Y=1) = 0.5$ sirvió como punto crítico de corte. La ventaja que presenta este método es que la comprobación de las hipótesis es fácilmente verificable. Los resultados permiten evidenciar que los factores identificados son estadísticamente significativos debido a que p -value es < 0.05 , es decir, son factores que influyen en la deserción estudiantil universitaria.

Pruebas: Después de la validación de los parámetros conviene efectuar pruebas de validación de los resultados para la comprobación de la calidad de la información muestral y la estabilidad de los datos frente al modelo especificado. En esta etapa también es importante señalar las divergencias si las hubiere, en relación con las hipótesis para determinar una especificación incorrecta del modelo.

Los residuos OLS dieron rangos positivos por encima de 0 y por debajo de 1, con lo cual se valida el supuesto de cumplimiento de $0 < E(Y_i | X_i) \leq 1$ y se verifica unos de los principios del modelo logístico (ver Figura 7).

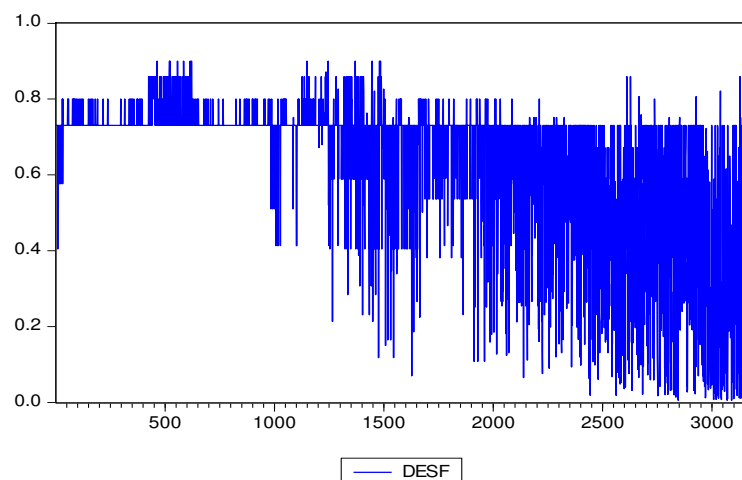


Figura 7. Residuos OLS del modelo de regresión logística
 Fuente: Elaborado por el autor

El contraste global del modelo fue evaluado por el índice de coeficientes de verosimilitudes o Pseudo-R² McFadden. El resultado presenta un valor 0.0133457, lo que evidencia que el modelo es estadísticamente significativo.

Por otro lado, para comparar las predicciones de las probabilidades de los datos observados, se utilizó el Test de Hosmer-Lemeshow, la medición de la bondad del modelo calculado ayuda a determinar si el modelo describe adecuadamente los datos, y se basa en las siguientes hipótesis:

H_{a0} = el modelo representa un buen ajuste de datos

H_{a1} = el modelo no representa un buen ajuste de datos

Tabla 29

Test Hosmer – Lemeshow

Goodness-of-Fit Evaluation for Binary Specification			
Andrews and Hosmer-Lemeshow Tests			
Grouping based upon predicted risk (randomize ties)			
		Prob.	Chi-
H-L Statistic	2.86	Sq(8)	0.000
Andrews Statistic	4.310	Prob. Sq(10)	Chi- 0.000

La estadística de Hosmer-Lemeshow indica un ajuste pobre del modelo, si el valor de la significación es mayor de 0.15. Con base a los resultados, se rechaza la hipótesis alternativa (H_{a1}) y se acepta la hipótesis nula (H_{a0}) en virtud de que el valor del Prob. Chi-Sq = 0.000, por lo tanto, se evidencia que el modelo propuesto presenta un buen ajuste de datos y se confirma que la variable (X) del modelo es la adecuada para estimar las probabilidades.

Para verificar el nivel de significancia y la estimación de los coeficientes de los parámetros del modelo propuesto, se utilizó el Test de Wald bajo los siguientes supuestos:

H_{b0} = X no influye en la deserción

H_{b1} = X influye en la deserción

Tabla 30
Wald Test

Wald Test:			
Equation: MODELO_LOGIT			
Test Statistic	Value	Df	P-value
F-statistic	3.030.830	(11, 3150)	0.000
Chi-square	3.333.913	11	0.000
Null Hypothesis:			
C(1)=C(2)=C(3)=C(4)=C(5)=C(6)=C(7)=C(8)=C(9)=C(10)=C(11)=0			

En vista de que $p\text{-value es } < 0.05$, se rechaza H_{b0} , es decir, todos los nuevos factores identificados influyen en la deserción estudiantil universitaria. Para determinar la probabilidad de la deserción estudiantil universitaria, se calculó el valor estimado de Y utilizando los coeficientes de la estimación por máxima verosimilitud, tal como se presenta en la tabla 31.

Se genera el proceso de simulación del modelo propuesto utilizando los valores de las estimaciones de (β), el valor de X y las estimaciones de Y. Para la variable X, se establece valores de ingreso de 1 a 3, por ejemplo, para el factor 1 $X=1$ (1 hijo), $X=2$ (2 hijos) y $X=3$ (3 hijos).

Tabla 31
Probabilidad de deserción estudiantil universitaria

Factor	(β)	Valor	Y(estimado)	P
F1	0.818043	X=1	0.8180	69%
		X=2	1.6360	84%
		X=3	2.2660	92%
F2	0.629028	X=1	0.6290	65%
		X=2	1.2580	78%
		X=3	1.8757	87%
F3	0.629028	X=1	0.6290	40%
		X=2	1.2580	31%
		X=3	1.8870	24%
F4	0.380309	X=1	0.3803	59%
		X=2	0.6814	68%
		X=3	1.1409	76%
F5	0.809717	X=1	0.8097	69%
		X=2	1.6194	83%
		X=3	2.4291	92%
F6	0.852620	X=1	0.8526	70%
		X=2	1.7052	85%
		X=3	2.5578	93%

Tabla 31
Probabilidad de deserción estudiantil universitaria (Continuación)

F7	0.687710	X=1	0.6877	67%
		X=2	1.3754	80%
		X=3	2.0631	89%
F8	1.384878	X=1	0.9999	73%
		X=2	1.9980	88%
		X=3	2.9970	95%
F9	0.281206	X=1	0.2812	57%
		X=2	0.5624	64%
		X=3	0.8436	70%
F10	0.637297	X=1	0.6372	65%
		X=2	1.2745	78%
		X=3	0.8712	87%
F11	1.346561	X=1	0.9990	73%
		X=2	1.9980	88%
		X=3	2.9970	95%

Dado que el modelo Logit es especificado como $Y(\text{estimado}) = \ln(p/1-p)$ y la probabilidad $(P) = \text{Exp}(Y \text{ estimado})/[1+\text{Exp}(Y \text{ estimado})]$, los coeficientes (β) se establecen como logaritmos de probabilidad, a fin de tomar el antilogaritmo o $\text{Exp}(\beta)$, se obtienen los coeficientes de $P/(1-P)$, esto significa que un incremento en la unidad de (β) incrementa la probabilidad de deserción estudiantil universitaria.

4.5 Discusión

Los resultados obtenidos en la presente investigación permitieron confirmar que los factores identificados son estadísticamente significativos para un nivel de confianza del 95 % y con un p-value del 0,05 %, por consiguiente, se pudo establecer que todos los nuevos factores influyen en la deserción estudiantil universitaria.

Los resultados de las estimaciones presentadas en la figura 8 permiten evidenciar que valores altos en los coeficientes (β) se relacionan con altas probabilidades deserción.

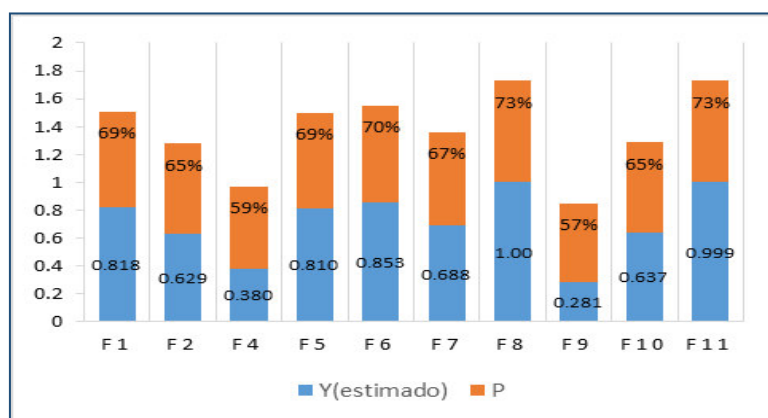


Figura 8. Probabilidad de deserción cuando $X=1$

Fuente: Elaborado por el autor

Los valores de las estimaciones de Y permitieron determinar que existe una relación de causalidad de los factores identificados con la deserción estudiantil universitaria, la simulación del modelo propuesto permitió evidenciar que existe una probabilidad de deserción superior al 57 %.

Los factores *número de hijos del estudiante (F8)* y *percepción del estudiante sobre la inserción del campo laboral (F11)* son los que presentan mayores probabilidades de deserción. Para el factor *número de hijos del estudiante*, el valor de $(\beta)=1$ indica que un incremento en la cantidad de hijos del estudiante en su etapa de formación universitaria aumenta la probabilidad de deserción, es así que si un estudiante tiene un hijo su probabilidad de deserción es del 73 %.

La alta influencia de este factor en el modelo puede deberse a que el 53 % de los estudiantes que contestaron la encuesta son mujeres, y el 11.13 % del total de encuestados reportaron tener al menos 1 hijo. Se puede considerar entonces que la presencia de hijos cuando los estudiantes están cursando un programa de estudios universitarios puede generar cambios en el estilo de vida que condicionan la probabilidad de finalizar con éxito los estudios en las universidades.

De igual manera, los resultados del factor *percepción del estudiante en la inserción del campo laboral (F11)* permiten establecer una probabilidad de deserción del 73 %. Se puede señalar que uno de los objetivos principales de las universidades es fortalecer las competencias académicas que permitan desarrollar las habilidades profesionales para que el estudiante pueda insertarse en el campo laboral. Si el alumno percibe que las condicionantes de formación universitarias no aportarán en la inserción del mercado laboral, este se verá motivado a buscar otra institución de educación superior que le brinde las facilidades académicas, tecnológicas y sociales para cumplir con sus expectativas laborales.

Por otro lado, en la figura 9 se puede evidenciar que el incremento del valor estimado Y (cuando $X=3$) genera un incremento en la probabilidad de deserción. Cuando los factores se activan individualmente el valor de *Probabilidad mínima* corresponde al 70 %, mientras la *Probabilidad máxima* alcanza el 95 %. Por consiguiente, se puede determinar que existe una alta relación de causalidad de los factores identificados y la deserción de los estudiantes en las universidades.

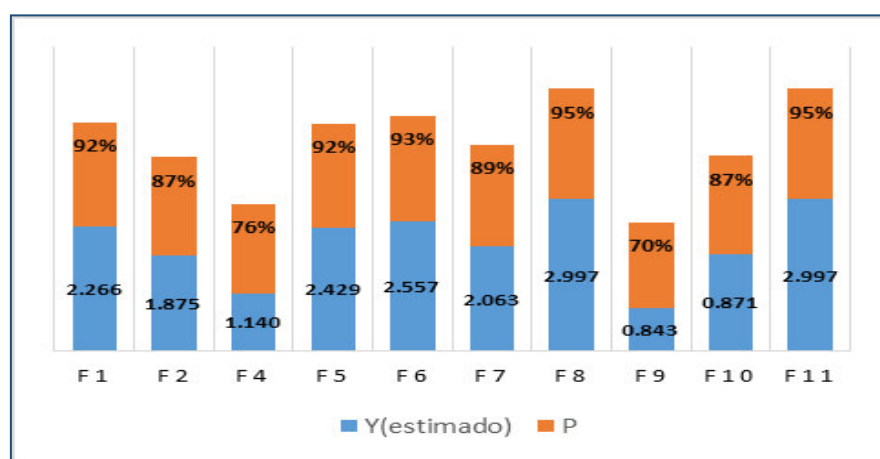


Figura 9. Probabilidad de deserción cuando $X=3$

Fuente: Elaborado por el autor

Por otro lado, en la figura 10 se puede identificar que el factor *compromiso del docente con el estudiante (F3)* presenta un valor de (β) con signo negativo. Los resultados indican que existe una relación inversa entre la

constante y el factor, es decir, la disminución del compromiso del docente en la formación académica y personal de sus estudiantes aumenta la probabilidad de deserción en un 40 %.

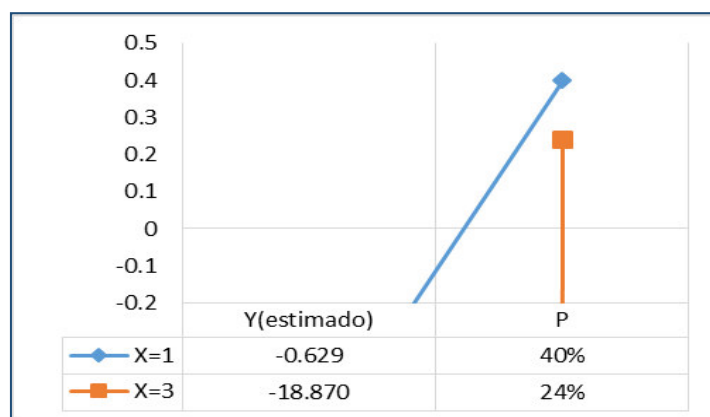


Figura 10. Probabilidad de deserción estudiantil universitaria

Fuentes: Elaborado por el autor

Finalmente, cuando todos los factores se activan a la vez con el menor valor ($X=1$), la probabilidad de que un estudiante deserte de sus estudios universitarios es del 95 %, con lo cual se pone evidencia la importancia de los factores identificados.

4.6 Conclusiones

La deserción en las universidades se encuentra relacionada con la influencia de varios factores que afectan negativamente en la decisión de los estudiantes de abandonar la universidad, bajo esta percepción el objetivo de este trabajo consistió en abordar la probabilidad de influencia que tienen los factores conocimiento limitado sobre el uso del software especializado de la carrera, embarazo planificado/no planificado, compromiso del docente con el estudiante, compromiso económico del hijo primogénito con la familia, bullying, machismo/feminismo, vicios adquiridos por el estudiante, número de hijos, adaptación del estudiante a las metodologías de formación universitaria, ranking de la carrera o institución, percepción del estudiante sobre la inserción en el campo laboral, con la deserción de los estudiantes

en las universidades. Con base a la metodología aplicada, se pudo comprobar que estos 11 nuevos factores identificados tienen una fuerte relación de causalidad con la deserción. Los factores identificados en su mayoría corresponden a la dimensión personal del estudiante, entonces se debería establecer estrategias que fomenten el vínculo estudiante-universidad en virtud de generar acciones de bienestar estudiantil que permitan al estudiante tener mayores posibilidades de culminar sus estudios universitarios y, por consiguiente, la reducción de las tasas de deserción. Los resultados presentados en la investigación proporcionan información valiosa para la toma de decisiones de los administradores de las Instituciones de Educación Superior.

CAPÍTULO V: PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL UNIVERSITARIA A TRAVÉS DE FACTORES TECNOLÓGICOS

5.1 Introducción

El abandono estudiantil universitario es considerado como un problema que afecta a las instituciones de educación superior en todo el mundo. Se ha convertido en un tema que genera polémica en el contexto educativo en el que están involucrados los administradores, profesores y estudiantes (Díaz Peralta, 2008). Se puede señalar que en la actualidad las tasas de deserción se analizan como criterios de calidad en los procesos de evaluación y acreditación de las universidades, que en muchos casos implica cambios académicos y sociales. (Djulovic & Li, 2013)

El interés que tienen los investigadores por tratar el problema de la deserción proviene de sus altas tasas y el elevado costo económico y social que afecta de manera negativa a los estudiantes y a las Instituciones de Educación Superior. En Chile, 3 de cada 10 estudiantes desertan antes o al finalizar el segundo semestre (Argote & Jiménez, 2014), mientras que en Estados Unidos, aproximadamente el 50 % de los estudiantes que ingresan en la universidad abandonan los estudios sin obtener un título profesional, problema que se evidencia más en las universidades públicas que en las privadas.

Por otro lado, la inclusión de la tecnología en la educación superior conlleva a la transformación del proceso educativo (Bustos Andreu & Nussbaum,

2009. Para Moreno y Molina (2014), los proyectos educativos que concretan las concepciones ideológicas de la enseñanza están influenciados a través de la ciencia, la tecnología y la educación; la sociedad del conocimiento en la universidad se vincula con el desarrollo de las competencias tecnológicas que exige en la actualidad el mundo globalizado. Según Henríquez y Escobar (2016), la educación de hoy está influenciada por la innovación de las Tecnologías de la Información y la Comunicación, el uso de estas tecnologías puede funcionar como mediación pedagógica que posibilite a los estudiantes encontrar situaciones que les permita al estudiante enfrentar la deserción universitaria Méndez –Estrada & Barrientos (2013).

Es importante recalcar que con el avance de la tecnología también se han incorporado factores tecnológicos que influyen en la decisión de los estudiantes de abandonar las universidades. Cuando los recursos tecnológicos como el internet, las redes sociales y la tecnología son utilizados de forma inadecuada, puede provocar en el estudiante efectos negativos que afectan en su proceso de permanencia de los estudiantes universitarios en las universidades. Para Blau (2011) y Ghamari, Mohammadbeigi, Mohammasdalehi & Hashiani (2011), el abuso del internet genera implicaciones negativas en el rendimiento educativo.

Los jóvenes adoptan tempranamente tecnologías digitales que los convierte en muchas ocasiones en usuarios sensibles a disturbios digitales como el abuso de las tecnologías (Gencer & Koc, 2012). Al tener los estudiantes universitarios mayor accesibilidad y uso del internet aumenta su vulnerabilidad al abuso de las tecnologías de información y comunicación debido a que tienen mayor libertad y largos periodos de tiempo no estructurados y estímulos otorgados por el entorno educativo para el uso del internet que conlleva en muchas ocasiones a la pérdida de la productividad educativa.

En el mismo sentido, Kittinger, Correia & Irons (2012) señalan que las redes sociales es una de las actividades populares relacionadas con el uso del

internet especialmente en estudiantes universitarios, el uso inadecuado de las redes sociales también es caracterizado como un comportamiento adictivo entre los jóvenes. Carbonell, *et.al.*, (2012) señala que existe relación entre la angustia psicológica y el uso inadecuado del internet y los teléfonos celulares que conllevan a trastornos que incluyen depresión, ansiedad e insomnio, lo que puede desorientar la misión original de los objetivos educativos.

A pesar de que el inventario de investigaciones relacionadas con el tema objeto de estudio es amplio, la mayoría de los trabajos analizados tiene como objetivo resaltar los beneficios positivos del uso de las tecnologías de información y comunicación en el proceso de enseñanza–aprendizaje universitario, dejando a un lado el análisis de la influencia negativa que puede ocasionar el inadecuado uso de las tecnologías de información y comunicación en la decisión de los estudiantes de abandonar las universidades. Teniendo en cuenta la necesidad de tomar medidas que permitan reducir el abandono estudiantil universitario, el presente artículo tiene como objetivo analizar la influencia que tienen los factores relacionados con el uso inadecuado del internet, redes sociales y tecnología en la deserción universitaria y evaluar a través de técnicas de minería de datos si estos factores pueden predecir el abandono escolar.

5.2 Metodología

El objetivo de la presente investigación es determinar el conjunto de factores tecnológicos para predecir la deserción estudiantil universitaria en las carreras de ingenierías en las universidades públicas del Ecuador, a través de un enfoque de aprendizaje automático basado en técnicas de regresión logística y árbol de decisión, para lo cual se establecen cuatro fases de la minería de datos que van desde el preprocesamiento de datos hasta la fase de pruebas, que se presenta en la figura 11.

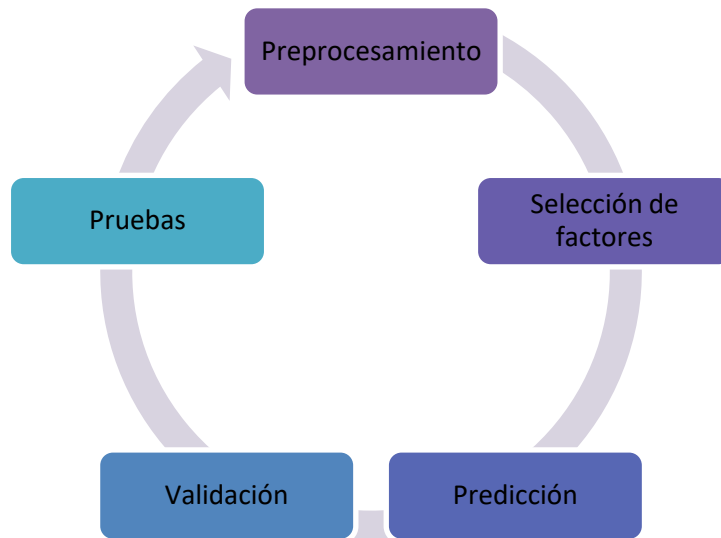


Figura 11. Proceso de predicción de la deserción universitaria
Fuentes: Elaborado por el autor

5.2.1 Conjunto de Datos

Se utilizaron datos de una encuesta realizada mediante Google Forms aplicada a 1178 estudiantes de las carreras de Computación, Electromecánica, Industrial y Eléctrica de varias universidades públicas de Ecuador. A partir de los objetivos de la investigación y la población a estudiar, se determinó el tipo de preguntas que deben aplicarse y sus características básicas, debido a que estas son consideradas como variables o indicadores respecto de los cuales se requiere obtener información. (Murillo, 2006)

La encuesta se dividió en tres secciones, la primera sección relacionada con las características de las instituciones y las personas que completaron el cuestionario, la segunda sección corresponde a preguntas que determinan los factores relacionados con el uso inadecuado de las tecnologías de información y comunicación en las universidades, y en la tercera sección las preguntas que complementan el estudio. Se realizó una prueba previa a 50 estudiantes antes del proceso de recopilación de datos, para verificar si las preguntas planteadas se encuentran relacionadas con los objetivos del estudio.

5.2.2 Factores tecnológicos

La tabla 32, presenta los factores tecnológicos considerados para el proceso de predicción de la deserción estudiantil universitaria. Se consideró la incorporación de los factores edad, género y estado civil junto con los factores tecnológicos que sirvieron como variables de ingreso a los modelos de predicción.

Tabla 32

Factores que influyen en la deserción estudiantil universitaria

Factor	Descripción
Género	Hombre, mujer
Edad	17 – 40 años
Estado civil	Casado, divorciado, viudo, unión libre
Internet	El estudiante usa el internet sin fines académicos
Redes sociales	El estudiante usa las redes sociales sin fines académicos
Tecnologías	Uso de las tecnologías (computadoras, celulares, tablet) sin fines académicos
Des	Intención del estudiante en abandonar los estudios universitarios

5.3 Resultados

El conjunto de datos incluye 7 factores que relacionan el uso inadecuado o abuso de las tecnologías de información y comunicación con el abandono de estudiantes en las universidades y que pueden constituirse en fuentes de variabilidad de las tasas de deserción. El factor deserción es considerado como la variable explicativa, mientras que los factores edad, género, estado civil, adicción al internet, adicción a las redes sociales y adicción a las tecnologías fueron considerados como variables explicativas.

Para la construcción de los modelos, se aplicaron algoritmos de minería de datos a través de dos técnicas árbol de decisión y regresión lógica. Se consideró la regresión logística, por ser un método de análisis para variables dicotómicas y porque presenta mayores estimaciones que otros métodos (Hosmer Jr, Lemeshow & Sturdivant, 2013). También, se utilizó el Decisión

Tree classifier, por ser una técnica de transformación para la conversión de las variables explicativas, la facilidad de interpretación de los resultados, y las mejores tasas de precisión (Márquez-Vera, Morales & Soto, 2013). Además, porque esta técnica utiliza una metodología que permite identificar las variables más significativas en el modelo y sus valores establecidos en una estructura en árbol. (Thammasiri, Denle, Meesad & Kasap, 2014)

Por otro lado, el inicio del proceso experimental empieza con el ajuste de los parámetros del conjunto de datos para ser utilizados en el aprendizaje supervisado. La etapa de preprocesamiento consta de tres actividades integración, limpieza y transformación de la información. Esta etapa fue utilizada para la gestión de anomalías, corrección de valores atípicos y valores perdidos, actividades que se realizaron para mejorar las propiedades de las variables, la optimización del proceso de búsqueda de los algoritmos de minería de datos.

En la etapa de selección de atributos relevantes se aplicó la técnica relief feature selection que evalúa el valor de un atributo al mostrar repetidamente una instancia considerando el valor del atributo dado para la instancia más cercana de la misma clase y diferente, el resultado del proceso de selección de factores se presenta en la tabla 33.

Tabla 33

Selección de factores de deserción estudiantil universitaria

Ranking de los atributos de los factores	
Attribute Evaluator (supervised, Class (numeric): 9 DES):	
	Relief Ranking Filter
	Instances sampled: all
Number of nearest neighbors (k): 10	
Equal influence nearest neighbors	
Ranked attributes:	
0.01017	6 Internet
0.00977	1 Gender
0.00844	7 Adic_Tec
0.00752	6 Red_S
0.00625	2 Age
-	3 Marital Status
0.00288	

Posteriormente, se estableció el proceso de predicción de la deserción estudiantil universitaria mediante la aplicación de los factores antes mencionados. Este proceso se utiliza para inferir atributos objetivos o aspectos de datos y en el caso de la educación es utilizada para detectar comportamiento de los estudiantes (Romero & Ventura, 2013). Además, se realizó 10 pruebas para cada algoritmo de Regresión logística, Decision Tree Classifier. A fin de comprender la importancia de las características de los factores utilizados como variables de ingreso a los modelos, se utilizó la métrica precisión para determinar la tasa de precisión de la predicción de la deserción. Los resultados permiten evidenciar que el método que presenta mayor tasa de precisión es Decisión Tree classifier con 91, 7 %, que se presenta en la figura 12.

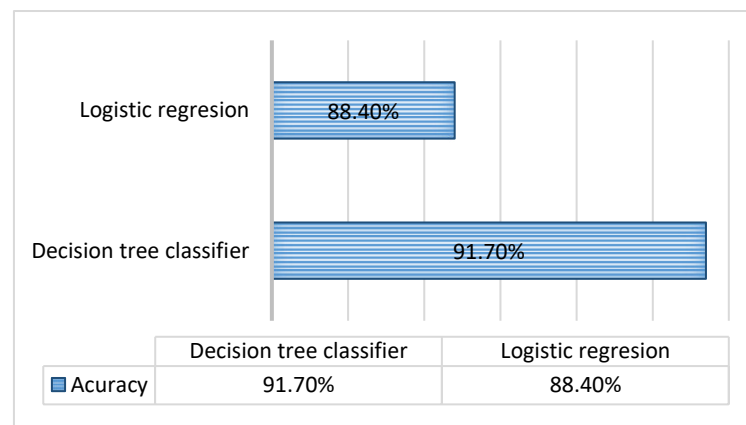


Figura 12. Precisión de los modelos de predicción

Fuente: Elaborado por el autor

5.4 Discusión y Conclusiones

El artículo presenta un enfoque de aprendizaje automático a través del uso de factores tecnológicos que influyen en la decisión de los estudiantes de abandonar la universidad. En este estudio se compararon dos técnicas de predicción árbol de decisión y regresión logística, junto con un conjunto de siete factores que sirvieron como variables de ingreso a los modelos de predicción. Para el proceso de predicción de la deserción estudiantil universitaria se estableció 4 etapas de la minería de datos que van desde la etapa de preprocesamiento hasta la etapa de predicción de la deserción.

Los resultados del estudio muestran que el método de árbol de decisión permitió predecir la deserción con un alto nivel de precisión correspondiente al 91.70 %. La determinación de los factores tecnológicos para predecir la deserción resalta la importancia del estudio debido a que se logró determinar la relación de causalidad de estos factores con la deserción en las universidades.

Con base en los resultados presentados en la tabla 2, se pudo establecer que el factor con mayor influencia, según el ranking de los atributos de los factores, fue la adicción al internet, que corresponde el uso del internet sin fines académicos con un peso de 0,1017. Esto puede deberse a que el 43.09 % de la población de estudiantes universitarios encuestada confirmó utilizar el internet sin fines académicos en un valor superior de 15 horas a la semana, lo que implica que los estudiantes universitarios usan esta cantidad de tiempo que podría ser utilizada con fines académicos.

A pesar de que el conjunto de datos utilizados en este estudio fue amplio, se considera que un incremento en el número de estudiantes encuestados y una mayor cantidad de factores podría mejorar los resultados de la predicción. Variables como el número de veces que los estudiantes universitarios ingresan a determinadas actividades del internet como facebook, instagram, wasap podrían ser consideradas como futuras investigaciones a partir de este estudio.

CAPÍTULO VI: ÁRBOL DE DECISIÓN PARA LA IDENTIFICACIÓN TEMPRANA DE ESTUDIANTES EN RIESGO DE DESERCIÓN

6.1 Introducción

La finalización de la universidad no siempre ha sido la norma para la sociedad, en la década de 1940, menos de la mitad de la población estadounidense comprendida entre edades de 25 a 29 años habrían terminado la universidad (Ye & Biswas, 2014). Si bien ha habido un esfuerzo concertado por cerrar la brecha de la deserción universitaria y disminuir sus tasas, investigaciones que parten desde 1978 evidencian que en las universidades todavía existen abandonados (Abu-Oda & El-Halees, 2015), lo cual ha causado efectos negativos en el orden económico para las instituciones de Educación Superior y los gobiernos.

La deserción estudiantil provoca dificultades en el contexto universitario (Archambault, Janosz, Dupéré, Brault & Andrew, 2017) y es considerado como un argumento de gran importancia dentro de la Ley de Educación a nivel mundial y en la incorporación de políticas públicas relacionadas con la educación universitaria (Hoyt, 1986). A pesar de que se han realizado cambios sistemáticos eficaces para resolver este problema, los estudiantes continúan enfrentando una crisis educativa con la mayor propensión a abandonar sus estudios. (López, 2004)

Además, la deserción genera consecuencias sociales en términos de las expectativas de los estudiantes y sus familias, así como también

consecuencias emocionales por la disonancia entre las aspiraciones de los jóvenes y sus logros. Las importantes consecuencias económicas, tanto para las personas como para el sistema en su conjunto, también son consideradas dentro de este contexto. (Vogel, Hochberg, Hackstein, Bockshecker, Bastiaens & Baumol, 2018)

La predicción de la deserción universitaria se torna importante desde hace algunas décadas, en donde se empieza analizar los factores que influyen en la deserción, combinado con la manera de predecir el riesgo de abandono (Alarcon & Edwards, 2013). Por otro lado, la predicción de la deserción en las Instituciones de Educación Superior ha sido cuestionada debido a las altas tasas de deserción estudiantil que aún tienen las instituciones. (Roso-Bas, Jiménez & García-Buades, 2016)

En la revisión de la literatura se encontraron trabajos como los de Márquez (2013), Herzog (2005), Kotsiantis, Zaharakis & Pintelas (2007), los autores establecen modelos de predicción de la deserción a través de procesos experimentales en los que se aplica métodos de aprendizaje automático supervisado para descubrir conocimiento.

De mantenerse los problemas antes expuestos, continuarán las altas tasas de error en la precisión de la predicción. Por tal razón, es importante establecer un modelo que permita integrar los datos, variables y técnicas adecuadas, para predecir con precisión los alumnos en riesgo de desertar. Además, permitirá a las Instituciones de Educación Superior contar con una herramienta eficaz para la toma acertada de decisiones en torno al tema de la deserción de los estudiantes universitarios.

6.2 TRABAJOS RELACIONADOS

La deserción estudiantil universitaria es un problema que ha sido ampliamente estudiado en la literatura. Se puede evidenciar los esfuerzos realizados por los investigadores en torno al tema de su predicción para tratar

de mitigar las tasas de deserción y establecer estrategias que permita la toma de decisiones de manera oportuna.

Varios trabajos se han identificado en torno al tema objeto de estudio y se presentan en la tabla 34.

Tabla 34

Trabajos relacionados con la aplicación de árbol de decisión para predecir la deserción en las universidades

Técnica	Fuente
Árbol de decisión	Yukselturk, Ozekes & Turel (2014), Hu, Lo & Shih (2014), Thammasiri, Delen, Meesad & Kasap (2014), Yasmin (2013), Bayer, Bydzovká, Géryk, Obsivac & Popellinsky (2012), Tan & Shao (2015), Aulck, Velagapudi, Blumenstock & West (2016), Xing, Chen, Stein & Marcinkowski (2016), Natek & Zwilling (2014), Lam-ON & Boongoen (2014), Li, Gao, Xiong, Wen & Wu (2016), Hoffait & Schyns (2017), Jadric, Garaca & Cukusic (2010), Pal (2012), Fei & Yeung (2015), Sultana, Khan & Abbas (2017), Sangodiah, Beleya, Muniandy, Heng & Ramendran (2015), Al-Barrak & Al-Razgan (2015), Guarín, Guzmán & González (2015), Heredia, Amaya & Barrientos (2015), Hershkovitz & Nachmias (2011), Liang, Yang, Wu, Li & Zhen (2016), Sivakumar, Venkataraman & Selvaraj (2016)

6.3 Materiales y métodos

6.3.1 Conjunto de Datos

Para el desarrollo de la investigación, se aplica una encuesta a 3162 estudiantes matriculados en modalidad de estudios presencial de pregrado de las Carreras de Ingeniería de una Universidad Pública del Ecuador. El período de análisis comprende las cohortes de estudio desde el año 2012 hasta el 2017. A través de la aplicación de Google Form, se obtuvo la información respecto al comportamiento de los estudiantes, específicamente información respecto a los hábitos de conducta de los alumnos, así se notó que puede generar

dependencia e influir negativamente en la decisión de abandonar las aulas universitarias.

6.3.2 Análisis de datos

La metodología aplicada para la aplicación de modelos de árbol de decisión se basa en las siguientes etapas:

- Etapa-1 Integración y limpieza de los datos: Realizada para obtener datos de calidad que permitan un proceso de predicción adecuado.
- Etapa-2 Preprocesamiento de los datos: Utilizada para determinar la normalidad y consistencia de los datos.
- Etapa-3 Extracción de variables: Utilizada para determinar las variables mayor incidencia en el modelo de predicción.
- Etapa-4 Predicción: Se aplica árbol de decisión para predecir la deserción de los estudiantes universitarios en función de las características de las variables de ingreso al modelo.
- Etapa-4 Evaluación: A través de métricas de evaluación, se determina el nivel de confiabilidad del modelo propuesto en términos de confianza y fiabilidad del modelo de predicción.

6.4 Resultados y Discusión

Árboles de decisión son técnicas que permiten la toma de decisiones basadas en el uso de probabilidades asociadas. Facilitan el nivel de comprensión respecto al comportamiento de las variables que influyen en la deserción de los estudiantes en las universidades. Las variables dependientes y las variables independientes utilizadas como variables predictores se presentan en la tabla 35.

Tabla 35
Descripción de las variables

Variables Independientes	Descripción
Red_S	Adicción a las redes sociales
Age	Edad
Alc	Adicción al alcohol
Adic_apegem	Adicción al apego emocional
Adic_drug	Adicción a las drogas
Adic_Tel	Adicción al usos del teléfono celular
Jueg	Adicción a los juegos
Videos	Adicción a los video juegos
Adic_com	Adicción a las compras
Variable dependiente	
DES	Deserción

La tabla 36 presenta el nivel de importancia de las variables independientes para el modelo de predicción. Los pesos de estas variables se determinaron utilizando el software SPSS.

Tabla 36
Importancia de las variables independientes

Variables independientes	Importancia	Importancia Standard
Red_S	.006	1.000
Age	.003	.444
Alc	.002	.403
Adic_apegem	.001	.096
Adic_drug	.001	.083
Adic_Tel	.000	.063
Jueg	.000	.006
Videos	.001	.096
Adic_com	.000	.002
Variable dependiente = (DES)		

La validación por división maestra corresponde al 80 % de los datos (2530 casos) para la muestra de entrenamiento y el 20 % (632 casos) para la muestra de comprobación. Se utiliza la validación cruzada 10 veces para evaluar los resultados del análisis estadístico y la validación del modelo propuesto.

Los criterios establecidos para el árbol de decisión CHAID corresponden a un número máximo de 3 niveles. El nivel de significancia para los nodos de división es igual a 0.05 al igual que la fusión de categorías.

Para el modelado del árbol de decisión, se consideró variables binarias, es decir:

0= los estudiantes desertarán de la universidad

1= los estudiantes no desertarán de la universidad

Para la evaluación del modelo, se consideró un número de estimaciones máximo de 100, así como un valor de cambio mínimo en las frecuencias esperadas de 0.001. El estadístico de Chi-cuadrado utilizado fue Person y se realizó la determinación de los valores de significancia a través del método de Bonferroni. El resumen del modelo propuesto se presenta en la tabla 37.

Tabla 37
Resumen del modelo

	Growth method	CHAID
	Dependent variable	DES
	Variables independientes	Age, Alc, Jueg, Video, Internet, Red_S, Adic_Tec, Adic_Com, Adic_apegem, Adic_Tel
Specifications	Validation	Split simple
	Maximum tree depth	3
	Minimum of cases in a subsidiary node	100
	Minimum of cases in a parent node	50
	Independent variables included	Red_S, Adic_Tel, edad, Adic_apegem, Adic_Com, Alc
Results	Number of nodes	15
	Number of terminal nodes	8
	Depth	3

La figura 13 presenta la estructura del árbol de decisión y la asignación de las probabilidades de los factores utilizados como variables de entrada al modelo.

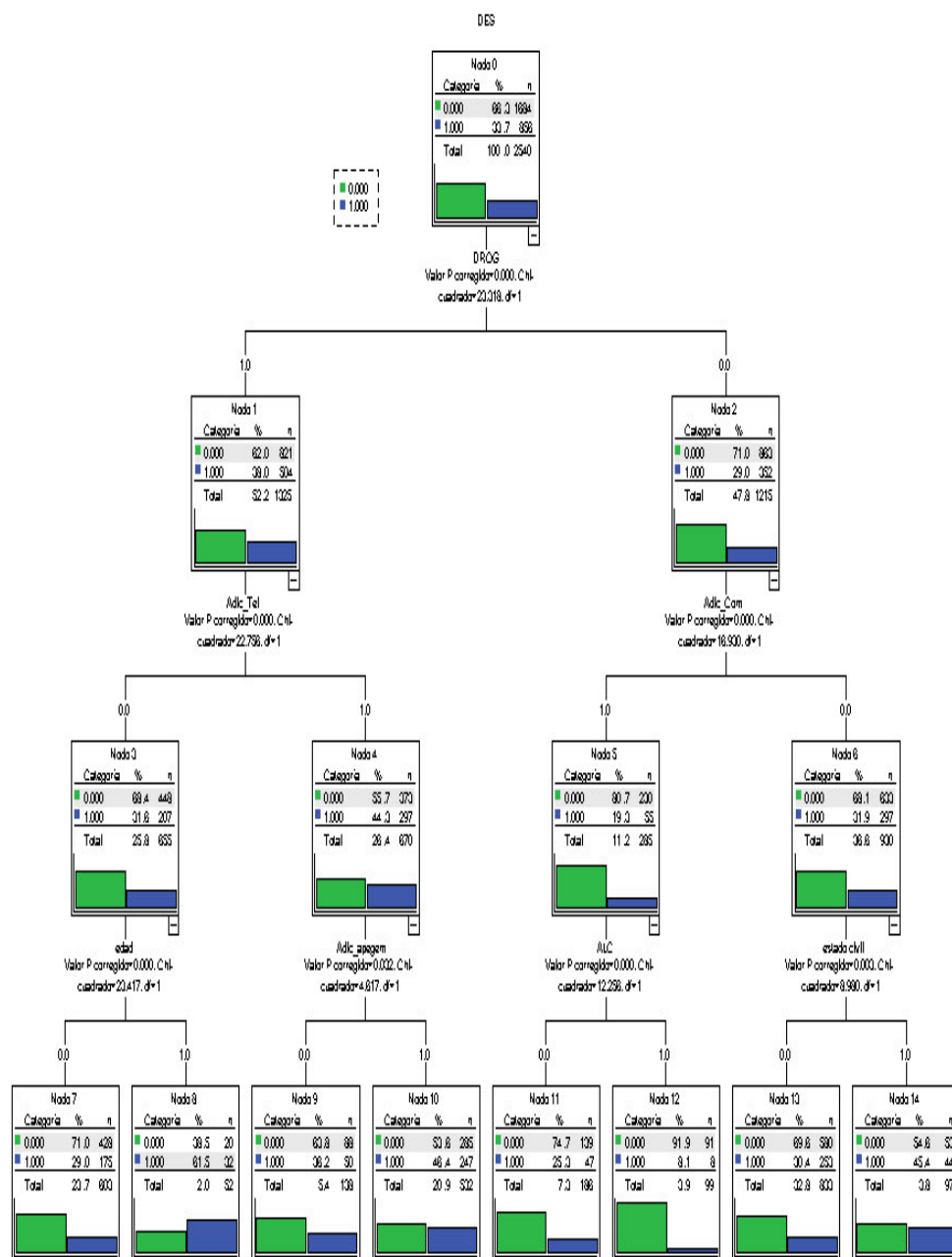


Figura 13. Árbol de decisión usando el método CHAID

Fuente: Elaborado por el autor

Una muestra de las reglas determinadas para predecir la deserción a través del método CHAIR se presenta a continuación.

/* Node 7 */.

DO IF (SYSMIS (Red_S) OR VALUE (Red_S) NE 0) AND (VALUE (Adic_Tel) EQ 0)
AND (SYSMIS (age) OR VALUE (age) NE 1).

COMPUTE nod_001 = 7.

COMPUTE pre_001 = 0.

```

COMPUTE prb_001 = 0.709784.
END IF.
EXECUTE.

/* Node 8 */.
DO IF (SYSMIS (Red_S) OR VALUE (Red_S) NE 0) AND (VALUE (Adic_Tel) EQ 0)
AND (VALUE (age) EQ 1).
COMPUTE nod_001 = 8.
COMPUTE pre_001 = 1.
COMPUTE prb_001 = 0.615385.
END IF.
EXECUTE.

/* Node 9 */.
DO IF (SYSMIS (Red_S) OR VALUE (Red_S) NE 0) AND (SYSMIS (Adic_Tel) OR
VALUE (Adic_Tel) NE 0) AND (VALUE (Adic_apegem) EQ 0).
COMPUTE nod_001 = 9.
COMPUTE pre_001 = 0.
COMPUTE prb_001 = 0.637681.
END IF.
EXECUTE.

```

Los resultados de los experimentos realizados muestran una tasa de precisión de la predicción correspondiente a 97.75 % para la muestra utilizada para el entrenamiento y del 98.71 % para la muestra de la prueba. Lo que significa que el modelo propuesto es adecuado en términos de calidad y efectividad.

Tabla 38
Resultados de la precisión del árbol de decisión CHAID

Sample	Observed	Predicted	
		0	1
Training	0	1664	20
	1	824	32
	Global percentage	97.95%	2.05%
Contrast	0	430	4
	1	184	4
	Global percentage	98.71%	1.29%

6.6 Conclusiones

La deserción de los estudiantes del Sistema educativo adquiere especial interés en las Instituciones de Educación Superior, especialmente en el sector público. Este problema causa efectos negativos como el fracaso del estudiante en el cumplimiento de sus objetivos académicos, generan también pérdidas

económicas para las instituciones y gobiernos de turno, la disminución en la tasa de graduación genera grandes problemas sociales e institucionales.

Los resultados del proceso experimental permitieron identificar que la variable adicción a las a las redes sociales, adicción al apego emocional, estado civil y edad fueron consideradas en el modelo propuesto como las variables de mayor influencia en el proceso de predicción.

Una tasa de precisión del 97.95 % permitió determinar que el método CHAID aplicado para predecir la deserción en las universidades es óptimo en términos calidad y efectividad. Se puede establecer que el modelo propuesto es eficaz para la identificación temprana de estudiantes en riesgo de abandonar las aulas universitarias. Además, puede ser utilizado por los administradores de las universidades como una herramienta de alerta temprana en la detección de estudiantes en riesgo de deserción. También puede convertirse en un instrumento de apoyo para la aplicación de políticas universitarias que permitan aumentar la tasa de permanencia estudiantil.

CAPÍTULO VII: RED NEURONAL PARA PREDECIR LA DESERCIÓN EN LAS UNIVERSIDADES

7.1 Introducción

La deserción estudiantil en las universidades es considerada como un problema que afecta a las instituciones de educación superior a nivel mundial (Sneyers & De Witte, 2017). En la actualidad las altas tasas de deserción son consideradas como posibles deficiencias en el sistema educativo de pregrado (Díaz, 2009). Esto se puede evidenciar en los informes académicos y de gestión administrativa que presentan las agencias gubernamentales a nivel mundial. En Estados Unidos la tasa de abandono del primer y segundo año alcanza el 44.8 % (Hoffait & Schyns, 2017), el 15.9 % corresponde a la tasa de deserción en la India (Wolf, David, Butler-Barnes & Zile-Tamsen, 2017), y en los países como Colombia, Ecuador y Brasil la deserción en las universidades supera el 40 % según la Organización de las Naciones Unidas (ONU) en el año 2016.

Como resultado, la deserción parece deberse a un conjunto de factores que interactúan entre sí y que ejercen influencia negativa sobre la decisión que toman los estudiantes para no alcanzar con éxito la culminación de sus estudios universitarios. (Enríquez y Escobar, 2016)

El abandono universitario es considerado como el abandono voluntario o involuntario de un programa de estudios (Himmel, 2018). Este fenómeno aún está presente en el sistema de educación superior y está relacionado con efectos negativos como el alto costo económico y social que afecta al estudiante en las universidades. (Eckert & Suénaga, 2015)

A pesar de que existe en la literatura una gran cantidad de estudios para tratar de resolver el problema de la deserción y determinar sus causas, en general existe limitada producción científica que incorpora algoritmos de machine learning como las redes neuronales para descubrir conocimiento en función de la naturaleza de las variables obtenidas a través del comportamiento de los estudiantes que permita determinar su condición de riesgo de deserción. Para diseñar acciones que permitan la disminución de las tasas de deserción en las universidades, se propone dos modelos de redes neuronales: perceptrón multicapa y función de base radial. Los resultados de esta investigación ayudarán a los administradores de las universidades a promover cambios en sus políticas y estrategias académicas con la finalidad de reducir sus tasas de deserción.

76.2 Trabajos relacionados

Varios estudios para predecir la deserción se han identificado en la revisión de la literatura, desarrollados para la identificación temprana de los estudiantes vulnerables y propensos abandonar las aulas universitarias. La tabla 39 presenta algunos trabajos que aplicaron en sus investigaciones redes neuronales para la predicción de la deserción en el contexto universitario.

Tabla 39

Redes neuronales para predecir la deserción en las universidades

Descripción	Referencias
<ul style="list-style-type: none"> • Predicción de la deserción en Turquía, basada en datos recopilados en un Programa de Tecnologías de la Información 	Yukselturk, Ozakes & Turel (2014)
<ul style="list-style-type: none"> • Sistema de alerta temprana para predecir el abandono escolar en las carrera de alfabetización Informal y Ética de la Información 	Hu, Lo & Shih (2014)
<ul style="list-style-type: none"> • Predicción de la deserción en España 	Lara, Lizcano, Martínez, Pazos & Riera (2014)
<ul style="list-style-type: none"> • Predicción de la deserción a través de características demográficas de los alumnos 	Thammasiri, Denle, Meesad& Kasap (2014)
<ul style="list-style-type: none"> • Predicción de la deserción en la República Checa a través de datos recogidos en estudiantes de la Carrera de Informática Aplicada 	Bayer, Bydzovská, Géryk, Obsivac & Popelinsky (2012)

Tabla 39

Red neuronal para predecir la deserción en las universidades (Continuación)

<ul style="list-style-type: none"> • Metodología para conectar el análisis de aprendizaje y la extracción de datos educativos y resolver el problema de deserción en la universidad 	Lykourantzou, Giannoukos, Nikolopoulos, Mpardis & Loumos (2009)
<ul style="list-style-type: none"> • Propuesta de análisis de clustering como marco de transformación de datos para identificar la predicción de deserción más precisa 	Lam-On, Boongoen (2014)
<ul style="list-style-type: none"> • Método de predicción del abandono escolar en la Universidad de México 	Dewan & Wen (2015)
<ul style="list-style-type: none"> • Predicción de la deserción en Europa a través de datos recogidos en estudiantes de las facultades de Ciencias Económicas 	Sultana, Khan & Abbas (2017)

7.3 Materiales y Métodos

7.3.1 Conjunto de datos

El conjunto de datos utilizado corresponde a información obtenida de 2670 estudiantes matriculados en modalidad de estudios presencial de pregrado de las Carreras de Ciencias Administrativas y Humanas de Universidades Públicas del Ecuador. Los datos fueron recolectados a través de una encuesta en línea, la misma que fue aplicada a estudiantes universitarios desde el primero al cuarto año académico. El período de análisis comprende las cohortes de estudio desde el año 2014 hasta el 2017.

La información de los estudiantes utilizados en la investigación comprende datos demográficos y del comportamiento de los estudiantes, información relacionada a las metodologías de formación universitaria, datos correspondientes a los procesos académicos e información socioeconómica de los alumnos.

7.3.2 Análisis de los datos

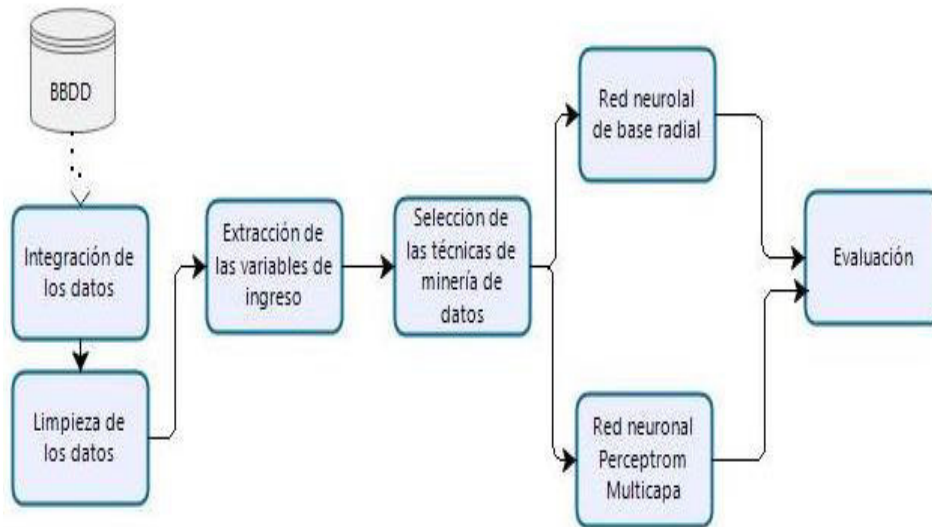


Figura 14. Proceso para predecir el abandono escolar a través de redes neuronales

Fuente: Elaborado por el autor

Como primer paso, se establece la integración y limpieza de los datos obtenidos de la encuesta, este paso se llevó a cabo para determinar que no exista redundancia de información, campos en blanco o datos que puedan afectar el proceso de predicción. El proceso de extracción de factores determina la importancia de las variables consideradas como predictoras en la capa de entrada de los modelos de red neuronal.

Para el proceso de modelado de datos, se seleccionaron dos técnicas de aprendizaje automático basadas en la función de base radial y perceptrón multicapa, que se utilizaron como un enfoque de aprendizaje automático para descubrir el conocimiento sobre el comportamiento de las variables relacionadas con abandono en las universidades.

El proceso de predicción se utilizó para determinar el grado de causalidad existente entre los factores de ingreso a los modelos y la deserción. Se llevó a cabo una evaluación de los modelos de redes neuronales a través de métricas de precisión de predicción como la sensibilidad.

7.4 Resultados y Discusión

7.4.1 Preprocesamiento de los datos

Se realizó la extracción de los factores más apropiados del conjunto de datos. Se seleccionaron mediante la determinación de ponderaciones que establecen la importancia de los factores en función de sus atributos para posteriormente aplicarlos al método de clasificación. La definición de las variables utilizadas se presenta en la tabla 40 y el valor correspondiente a los pesos de las variables se presenta en la tabla 3.

Tabla 40
Definición de variables

ID	Factor
F01	Conocimiento limitado sobre el uso de software especializado de la carrera
F02	Embarazo planificado/no planificado
F03	Compromiso del profesor con el alumno
F04	Compromiso económico con la familia por ser hijo primogénito
F05	Bullying
F06	Machismo/feminismo
F07	Adicciones del estudiante
F08	Número de hijos del estudiante
F09	Adaptación del estudiante a las metodologías de aprendizaje
F10	Ranking de la Universidad o Carrera
F11	Perspectiva del estudiante sobre la integración en el campo laboral
DES	Deserción

Tabla 41
Importancia de las variables dependientes

Radial base function			Perceptrón multilayer	
Id	Importancia	Importancia Standard	Importancia	Importancia Standard
F3	0,090	0.520	0.0827	0.653
F7	0.080	0.465	0.100	0.796
F1	0.173	1	0.108	0.858
F4	0.104	0.599	0.105	0.830
F10	0.103	0.593	0.078	0.620
F2	0.058	0.339	0.088	0.695
F8	0.093	0.538	0.126	1
F5	0.076	0.438	0.066	0.524
F11	0.060	0.346	0.105	0.837
F9	0.069	0.398	0.0723	0.571
F6	0.088	0.510	0.064	0.512

7.4.2 Predicción de la deserción estudiantil universitaria

El objetivo del proceso experimental es demostrar que las redes neuronales se pueden usar como algoritmos de alto rendimiento para predecir el abandono escolar. Por lo tanto, se realizó una comparación entre dos algoritmos: perceptrón multicapa y función de base radial. En segundo lugar, se evalúa el rendimiento de métodos aplicados para determinar las diferencias significativas de los resultados obtenidos en términos en precisión de predicción.

Para Montaña (2017), las redes neuronales se consideran como sistemas de procesamiento de información basados en conexiones a través de neuronas. Funciona a través de la suma de los pesos de los factores que se comparan con los valores de umbral para obtener una activación de la neurona de salida. Su objetivo es seleccionar, en función de los parámetros internos de la capa de entrada, las relaciones entre las conexiones de las neuronas a través del entrenamiento de la red (Torres, Ramos & Moraga, 2016). Las redes neuronales son uno de los métodos más populares de aprendizaje automático. Se utiliza ampliamente en la educación por su capacidad para reconocer patrones en el comportamiento de los estudiantes basados en los factores identificados. La suma ponderada de los valores de entrada a la red neuronal se consideró con base en los pesos asignados a cada de las funciones de entrada presentados en la ecuación (1), y a su vez es utilizada para definir las redes neuronales de la red y considera como parámetros de entrada del modelo propuesto.

$$F = (a_1x_1 + a_2x_2 + a_3x_3 \dots \dots + b + a_nx_n) \quad (1)$$

En donde:

a= al peso de la variable

x= variable de ingreso a red neuronal

y= función salida de la neuronal

b= sesgo

Para modelar la red neuronal, se considerarán las variables binarias, es decir 0 para predecir que los estudiantes abandonarán la universidad y 1 para predecir que los estudiantes universitarios no abandonaran la universidad.

Por otro lado, para la construcción del primer modelo, se utilizó perceptrón multicapa. Los valores de los pesos se determinan a partir de la muestra de entrenamiento y la función de error de la red neuronal considerada como el error cuadrático medio (Al-Omari & Bousa, 2014). Las opciones de aprendizaje para la red neuronal perceptrón multicapa se presentan a continuación.

```
MLP des (mlevel=n) by F3 F7 F1 F4 F10 F2 F8 F5 F11 F9 F6
Partition training=6 testing=3 holdout=1
Architecture automatic=yes (minunits=1 maxunits=50)
Criteria training=batch optimization=scaledconjugate lambdainitial=0.0000005
sigmainitial=0.00005 intervalcenter=0 intervaloffset=0.5 memsize=1000
Print cps networkinfo summary classification solution importance
Stoppingrules errorsteps= 1 (data=auto) trainingtimer=on (maxtime=15)
maxepochs= auto errorchange= 1.0e-4 errorratio=0.0010
```

El conjunto de datos de partición corresponde al 60 % (1602 casos) de entrenamiento y el 30 % (801 casos) para la prueba, dejando un 10 % (267 casos) para la reserva.

En la arquitectura del modelo, se consideró como número mínimo de unidades de la capa oculta 1 y un valor máximo de unidades de capa oculta 50. Para el proceso de entrenamiento de la red neuronal, se utilizó la función logarítmica sigmoide. Además, se utilizó la entropía cruzada como función de error.

El número de capas ocultas 1^a es igual 9 y la función de activación de la capa oculta utilizada fue la tangente hiperbólica, mientras que la función de activación de la capa de salida fue softmax. Además, se utilizó como función de error la entropía cruzada. El resumen del modelo utilizado se presenta en la tabla 42. Los resultados obtenidos de la red neuronal se obtuvieron mediante la aplicación del software SPSS.

Tabla 42

Resumen del modelo utilizando el método Perceptrón Multicapa

Training	Cross entropy error.	1166.316
	Percentage of incorrect forecasts	0.126
	Stop rule used	1 pasos consecutivos sin disminución del error ^a
	Training time	00:00:01,187
Testing	Cross entropy error.	566.285
	Percentage of incorrect forecasts	0.118
Backup	Percentage of incorrect forecasts	0.134

La figura 15, presenta la curva de elevación de la variable dependiente (DES). Para el modelo de predicción que usa perceptrón multicapa.

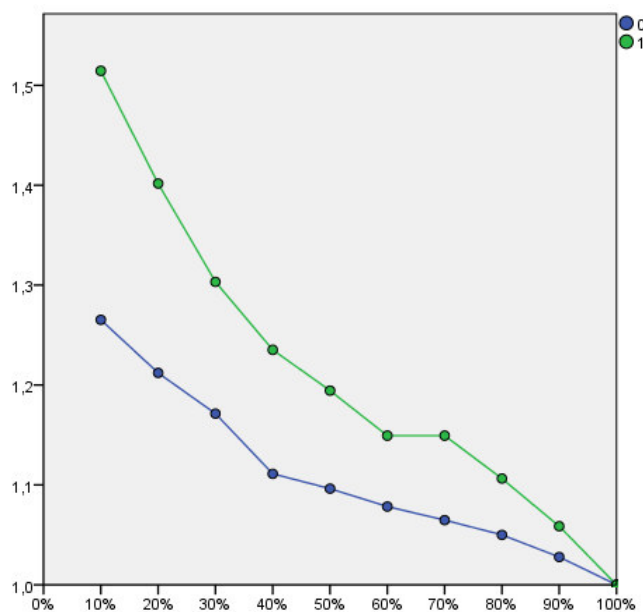


Figura 15. Curva de elevación variable dependiente (DES), modelo perceptrón multicapa

Fuente: Elaborado por el autor

Respecto al segundo modelo, se utilizó la función de base radial. El conjunto de partición de datos con el 70 % (1869 casos) de entrenamiento y el 40 % (801 casos) para la prueba. Las opciones de formación de la red neuronal se presentan a continuación:

Radial basis function network.
 RBF des (mlevel=n) by F3 F7 F1 F4 F10 F2 F8 F5 F11 F9 F6
 Partition training=7 testing=3 holdout=0
 Architecture minunits=auto maxunits=auto hiddenfunction=nrbf
 Criteria overlap=auto
 Print cps networkinfo summary classification importance
 Missing usermissing=exclude.

11 factores como variables independientes para la capa de entrada. En la capa oculta, el número de unidades es igual a 7^a mediante la función de activación softmax.

Para la capa de salida, se consideró nuevamente la variable dependiente (DES) se consideró nuevamente. El número de unidades es igual a 2 mediante la función de activación identidad y la función de error de suma de cuadrados. El número de unidades ocultas está determinado por el criterio de datos de prueba, el número óptimo de unidades ocultas es el que produce el menor error en los datos de prueba. El resumen del modelo utilizado se presenta en la tabla 43.

Tabla 43
Resumen del modelo utilizando el método Función de Base Radial

Training	Sum of quadratic errors	485.548
	Percentage of incorrect forecasts	0.116
	Training time	00:00:02,613
Testing	Sum of quadratic errors	210.722 ^a
	Percentage of incorrect forecasts	0.105

La figura 16, presenta la curva de elevación de la variable dependiente (DES) para el modelo de predicción que utiliza función de base radial.

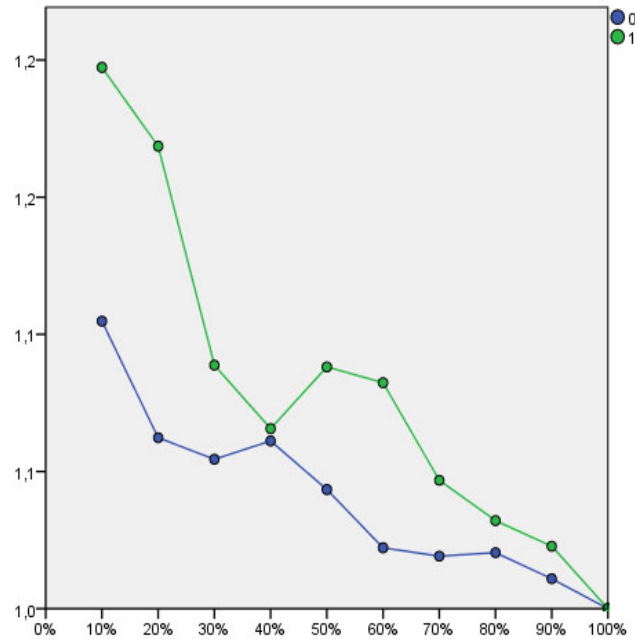


Figura 16. Curva de elevación variable dependiente (DES), modelo función de base radial

Fuente: Elaborado por el autor

Por otro lado, para evaluar la viabilidad de los modelos propuestos se analiza la curva ROC (características de funcionamiento del receptor) y se determina la capacidad predictiva de los modelos de predicción. Este tipo de gráficos se crean con base en las pseudoprobabilidades que utilizan el error de las sumas cuadráticas y la función de activación de la capa de salida. La figura 17 muestra la representación visual de la sensibilidad determinada en la columna (Y) frente a la especificidad presentada en la columna (x) en relación con la variable dependiente (DES). La sensibilidad métrica se considera como la relación de verdaderos positivos, mientras que la especificidad es considerada como la relación de verdaderos negativos de los puntos de corte de la red neuronal. (Witten, Frank, Hall & Pal, 2016)

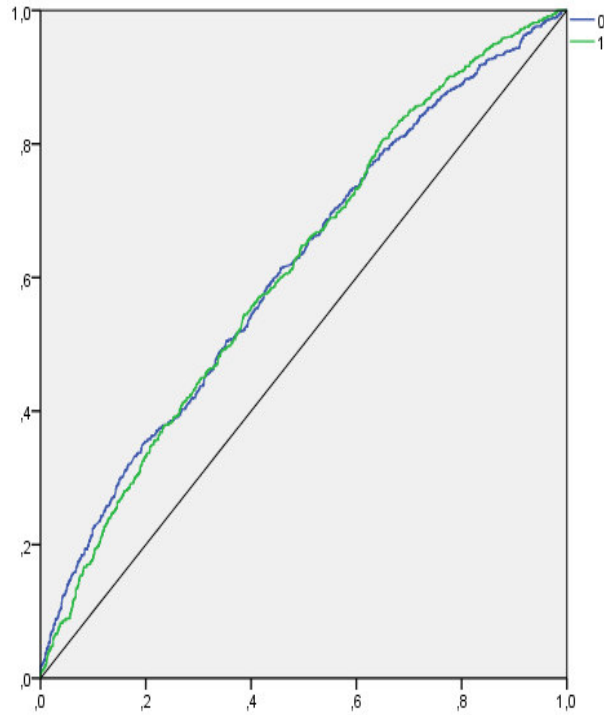


Figura 17. Sensibilidad (Y) vs Especificidad (x)

Fuente: Elaborado por el autor

Como se puede observar, ambas técnicas muestran un rendimiento óptimo en términos de precisión de predicción en las universidades. No hay diferencias significativas en cuanto a la predicción de los dos modelos, por lo tanto, se puede considerar que estos pueden ser una opción eficaz para predecir con acertado grado de precisión los estudiantes en riesgo de abandonar las aulas universitarias. Los resultados del proceso de predicción se presentan en la figura 18 y en la tabla 44.

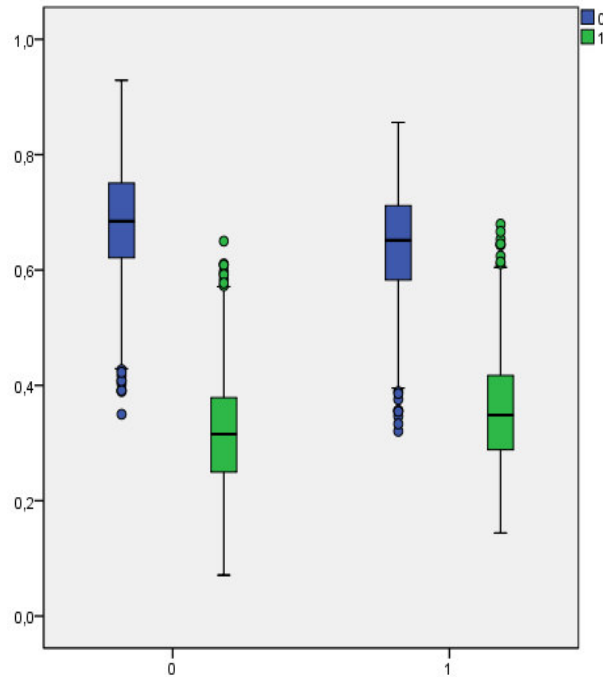


Figura 18. Pseudoprobabilidad predicha

Fuente: Elaborado por el autor

Tabla 44

Precisión de la predicción de los modelos propuestos

Redes Neuronales	Entrenamiento	Prueba
Perceptrón multicapa	96,3%	98,6%
Function de Base Radial	96,8%	98,1%

7.5 Conclusiones

El artículo presenta una aplicación de redes neuronales mediante la implementación de los algoritmos perceptrón multicapa y función de base radial como métodos de predicción de la deserción estudiantil universitaria.

Los resultados de los análisis mostraron que perceptrón multicapa presente una tasa de predicción del 96.3 %, mientras que la tasa de precisión de la función base radial corresponde al 96.8 %. Lo que indica que los modelos de redes neuronales son fiables para la identificación de los estudiantes en riesgo de desertar de las universidades.

Estos resultados pueden ayudar a los administradores de las universidades en la toma de decisiones oportuna y la implementación de estrategias que permita la reducción de sus tasas de deserción.

CAPÍTULO VIII: METACLASIFICADOR PARA PREDECIR LA DESERCIÓN ESTUDIANTIL UNIVERSITARIA

8.1 Introducción

La deserción es considerada como el abandono prematuro de manera voluntaria o involuntaria de un programa de estudios, en el cual el alumno descarta la finalización exitosa de sus estudios (Himmel, 2018). Para Pereira, Romero & Toledo (2013), la deserción es considerada como la situación a la que se enfrenta un estudiante cuando aspira y no logra finalizar su proyecto educativo. Este fenómeno se mantiene en la actualidad presente en el sistema de educación superior y está relacionado con efectos negativos como el elevado costo económico y social que afecta al estudiante y a las universidades. (Eckert & Suénaga, 2015)

Las altas tasas de deserción han generado impacto en las universidades desde hace varios años, Larsen. *et al.* (2013), que comprometen la estabilidad de ingresos de las universidades y la calidad de procesos académicos y administrativos, debido a que el costo económico de la educación universitaria en el sector público en muchos países es asumido por las universidades y los estados, Gartner Isaza, Dussán Lubert & Montoya (2016).

Alrededor del 17-18 % de los estudiantes europeos abandonan sus estudios antes de graduarse (Roso, Jiménez & García, 2016). Según Elias (2008), la tasa media de abandono de los estudiantes en las universidades españolas comprende el 26.7 %. Mientras que en Italia solo un tercio de los estudiantes inscritos en los programas universitarios obtienen un título universitario. (Gitto, Minerviri & Monaco, 2016)

En Estados Unidos, aproximadamente, una cuarta parte de los estudiantes universitarios abandonan sus estudios al final del primer año (Xing, Chen, Stein, Marcinkowski, 2016), en Noruega, más de la mitad de los estudiantes universitarios abandonan la universidad antes de la graduación, Simpson (2013).

El análisis de la deserción permite identificar la importancia que tienen la determinación de factores académicos, socioeconómicos y personales que influyen en este problema, Lozano (2018). Entonces, se considera que la deserción puede ser de naturaleza multifactorial, es decir, puede estar involucrada la influencia de múltiples factores en la decisión del estudiante de abandonar las aulas universitarias, Stewart, Lim & Kim (2015) y Roso, Jiménez & Gracia (2016).

Por otro lado, el acelerado desarrollo en el escenario universitario plantea nuevos retos para los investigadores respecto al análisis y revisión de las trayectorias de la predicción de la deserción. Se identifica varios estudios para predecir el abandono universitario en la revisión de la literatura, sin embargo, la predicción de la deserción en las universidades puede variar dependiendo de las variables de ingreso a los modelos, del contexto educativo estudiado, del entorno de educación aplicado y de los antecedentes de los estudios para los que fueron aplicados y del tamaño del data set.

Uno de los principales métodos de análisis que se aplica para este propósito es la minería de datos, que se define como el proceso no trivial de la identificación de patrones válidos. Al considerar la necesidad de diseñar acciones que propendan la disminución de abandonos en las universidades, este trabajo tiene como propósito diseñar un modelo híbrido para predecir la deserción estudiantil universitaria. El modelo propuesto considera 5 técnicas de minería de datos y 11 nuevos factores de deserción que sirvieron como variables predictores de ingreso al modelo propuesto.

8.2 Materiales y métodos

8.2.1 Conjunto de datos

El conjunto de datos utilizado para el análisis se construye de una encuesta en línea a través de Google Form. Se aplicó la encuesta a 6000 estudiantes universitarios de carreras universitarias de Ciencias Técnicas, Administrativas, Humanísticas, en la modalidad presencial, en el Ecuador. El objetivo de la encuesta fue determinar la percepción que tienen los estudiantes universitarios sobre factores que influyen en la deserción y la percepción de los 130 expertos con experiencia en el tema objeto de estudio.

La encuesta se estructuró en tres secciones. En la primera sección se establecieron preguntas relacionadas con las características de las universidades donde estudian. La segunda sección se relacionó con los factores que influyen en la deserción. En la tercera sección se estableció preguntas que complementaron el estudio.

Una vez completada la encuesta, se llevó a cabo una prueba piloto a 200 estudiantes para verificar si el idioma era comprensible y si las preguntas estaban claramente establecidas. Como resultado se obtuvieron 11 factores que influyen en la predicción de la deserción, los que fueron utilizados como variables de ingreso al modelo híbrido propuesto.

8.2.2 Análisis de los datos

La tabla 45 presenta las estadísticas descriptivas de los factores de los estudiantes encuestados y que fueron considerados como sujetos de estudio.

Tabla 45

Estadística descriptiva de los factores identificados

Factor	Mean	Std. Dev.	Std. Err. of Mean
--------	------	-----------	-------------------

Dropout (DES)	0.330171	0.470349	0.008364
Limited knowledge about specialized software usage in the university career	4.273.245	0.889305	0.015815
Planned/unplanned pregnancy	4.283.049	1.003.895	0.017853
Teacher's commitment towards the student	4.166.983	1.085.091	0.019297
First son's economic commitment towards his family	4.353.257	1.041.744	0.018526
Bullying	4.021.822	0.913794	0.016251
Sexism	4.032.258	1.021.860	0.018172
Student's acquired addictions	4.053.447	0.8918	0.015859
Student's number of children	4.246.996	1.078.110	0.019173
Student's adaptation to the university learning	4.237.192	1.172.837	0.020857
Career or institution ranking	4.206.831	1.137.597	0.020231
Student's perspectives towards the integration into the labor market	4.332.385	0.990889	0.017622

8.2.3 Análisis de los métodos

Los modelos de aprendizaje automático utilizados para predecir la deserción estudiantil universitaria se describen a continuación:

Árbol de decisión: La técnica de clasificación y predicción es muy utilizada por sus múltiples ventajas como un rápido proceso clasificación de registros desconocidos, fácil interpretación y estructura robusta para tratar valores atípicos (Yukselturk, Ozeques y Türel, 2014). También es considerado como un predictor preponderante en el tratamiento de variables categóricas y cuando se requiere explicación de una variable de respuesta individual y de datos de grupos homogéneos para la obtención de patrones. (Yasmin, 2013)

Redes Neuronales: Es un conjunto de variables interconectadas de entrada/salida y cada conexión tiene un peso presente, durante la fase de aprendizaje la red aprende mediante el ajuste de pesos a fin de poder predecir la clase de etiquetas correctas de las tuplas de entrada (Suhirman, 2014). Las redes neuronales tienen la capacidad de entender el significado de los datos complicados e imprecisos y son muy adecuadas para las necesidades de predicción y pronóstico.

Redes neuronales: Están establecidas para encontrar hallazgos de elementos frecuentes entre grandes conjuntos de datos (Romero, 2010). Son capaces de generar reglas con confianza de valores inferiores a 1 y es muy utilizada en la minería de datos en el análisis de correlaciones y en procesos de toma de decisiones.

8.3 Resultados y discusión

8.3.1 Preprocesamiento de los datos

Permite el manejo de anomalías, corrección de valores atípicos y valores faltantes (Stratton, 2008). Su propósito es mejorar las propiedades de las variables y resolver problemas para optimizar el proceso de búsqueda de los algoritmos de minería de datos. Estas técnicas están relacionadas con tres actividades importantes: integración, limpieza y transformación de la información. La técnica utilizada para este propósito fue la Normalización, en donde el atributo es escalado a un rango específico de -1 a 1 y de 0 a 1, el proceso de normalización se presenta en la ecuación (1).

$$V' = \frac{v - \min A}{\max A - \min A} (\text{newmax}A - \text{newmin}A) + \text{newmin}A \quad (1)$$

8.3.2 Extracción de factores

Para la extracción de factores, se aplica el algoritmo CfsSubsetEval, que evalúa el valor de un subconjunto de atributos, donde se considera la capacidad predictiva individual de cada característica junto con el grado de redundancia entre ellos. Mediante el método GreedyStepwise, se realiza una codiciosa búsqueda de factores con los mejores atributos hacia delante o hacia atrás a través del espacio de subconjuntos de atributos.

Tabla 46
Selección de atributos

=== Attribute Selection on all input data ===
Search Method:

Attribute ranking.		
Attribute Evaluator (supervised, Class (nominal): 15 DES):		
ReliefF Ranking Filter		
Instances sampled: all		
Number of nearest neighbours (k): 10		
Equal influence nearest neighbours		
Peso	ID	Factor
0.152	F9	Adaptación del estudiante a las metodologías de aprendizaje universitario
0.137	F1	Limitado conocimiento en tecnologías especializadas de la carrera
0.133	F12	Género
0.126	F7	Adicciones del estudiante
0.124	F10	Ranking de la universidad
0.114	F8	Número de hijos
0.105	F5	Bullying
0.105	F6	Machismo/feminsimo
0.104	F2	Embarazo planificado/no planificado
0.100	F11	Perspectiva del estudiante en la integración en el mercado laboral
0.099	F4	Compromiso económico del estudiante
0.096	F3	Compromiso del docente con el estudiante
0.095	F13	Estado civil
0.091	F14	Edad

8.3.3 Predicción de la deserción

Esta sección presenta los resultados de la predicción de la deserción estudiantil universitaria en función de las características de 11 factores que influyen en la deserción del estudiante que es el actor principal de la deserción. El conjunto de datos fue utilizado para las fases entrenamiento y pruebas de los algoritmos seleccionados. En este estudio se utilizó métodos de aprendizaje supervisado, redes neuronales, support vector machine, naive bayes, árbol de decisión con J48 y reglas de clasificación mediante el clasificador PART, donde se utiliza la validación cruzada para cada clasificador. Además, se completa diez pruebas por cada algoritmo y se aplica métricas de rendimiento para comparar la precisión de los modelos.

Los valores de las salidas de los modelos dieron como resultados datos binarios que corresponden a 0 con una estimación de que el estudiante

completará sus estudios universitarios y el valor de 1 que corresponde a la probabilidad de riesgo de deserción.

Tabla 47
Porcentaje de precisión de la predicción

Técnica	TP Rate	FP Rate	Recall	F-Measure
J48	0,988	0,020	0,988	0,988
PART	0,994	0,027	0,994	0,990
Random Forest	0,986	0,020	0,986	0,986
Multilayer perceptron	0,980	0,032	0,980	0,980

El objetivo de la investigación fue predecir el riesgo de deserción de los estudiantes en las universidades. El valor de F-measure superior a 0,98 determina la eficacia general de los modelos de predicción de la deserción estudiantil universitaria. Se puede determinar que los valores obtenidos son relativamente eficaces en la identificación del riesgo de abandono de los estudiantes en la clase deserción.

Respecto a la precisión de las técnicas de minería de datos, se consideró la medida de rendimiento accuracy, que representa la precisión de un instrumento (Lara, 2014) y es dado por (1).

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Donde TP es la cantidad de verdaderos positivos, TN la cantidad de verdaderos negativos, FP la cantidad de falsos positivos y FN la cantidad de falsos negativos. Los parámetros de precisión de clasificación de datos fueron medidos por la herramienta Weka, los resultados se presentan en la Figura 18.

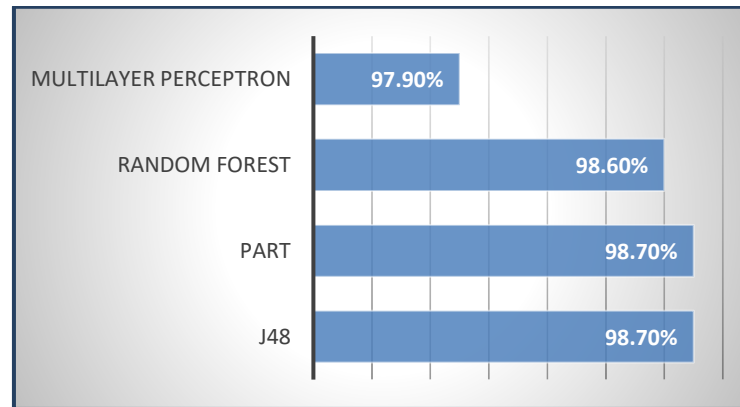


Figura 18. Precisión de los modelos de predicción

Fuente: Elaborado por el autor

Con base en los resultados presentados, se puede determinar que existe poca diferencia en términos de tasa de rendimiento de los modelos de predicción. Los algoritmos PART y J48 presentan resultados de precisión más precisa (98.7 %) aunque los dos modelos tienen la misma tasa de éxito. El valor de F-measure (0.990) del primer algoritmo es superior al otro, el mismo que posee un porcentaje del 97.90 %.

Durante el entrenamiento, la técnica de aprendizaje ajusta sus parámetros internos para deducir la asignación implícita de los datos de entrenamiento proporcionados. Para la fase de prueba, cada técnica se presenta con datos que no se han utilizado para el entrenamiento para examinar el rendimiento de los clasificadores. Las decisiones tomadas de los cinco clasificadores entrenados se combinan en paralelo usando el meta clasificador votated, que es un meta algoritmo que permite integrar los algoritmos de aprendizaje supervisado para mejorar la tasa de predicción de la deserción. El resultado obtenido se presenta en la figura 19.

Al igual que en el resto de algoritmos, se utilizó la validación cruzada 10 veces para el meta clasificador. El conjunto de reglas se asocia a los atributos de los factores que se utilizaron como variables de ingreso a los modelos de predicción de la deserción.

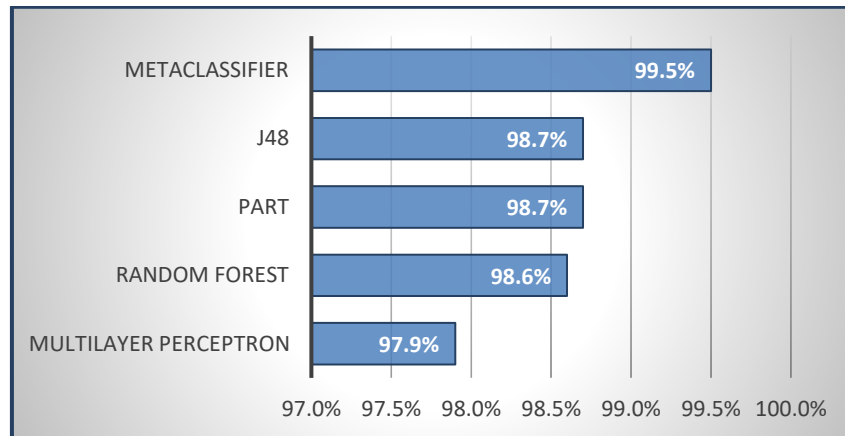


Figura 19. Precisión del meta clasificador

Fuente: Elaborado por el autor

Para mejorar la precisión de predicción, se utilizó el meta clasificador vote, a través de la combinación de clasificadores de alto rendimiento como son Multilayer Perceptrón, Random Forest, PART y J48, donde se utilizó el software WEKA. Uno de los aspectos considerados para la construcción del meta clasificador fue el número de clasificadores de alto rendimiento. Se combinó las salidas de los clasificadores a través del método combination rule maximun probabilities, los resultados se presentan en la figura 2.

Los resultados evidencian que al combinar los clasificadores mediante un método de fusión, la precisión aumenta respecto al desempeño de los clasificadores ejecutados de forma individual. Existe diferencia significativa entre los clasificadores multilayer perceptron, random forest, PART y J48. A través del metaclasificador Vote, se mejora el porcentaje de la presión de la deserción, donde su valor llega a 99.5 %.

Los experimentos ejecutados dieron como resultado valores significativos respecto a las medidas de concordancia que permitieron evaluar el rendimiento del meta clasificador. El coeficiente kappa statistic toma valores entre 0 y 1, donde valores superiores a 0.75 pueden considerarse como un grado óptimo de precisión y fiabilidad del modelo. Los resultados del proceso experimental presentan una estimación puntual de los intervalos de confianza de kappa statistic del 0.982, lo que indica el modelo de predicción de la

deserción estudiantil universitaria propuesto es óptimo en términos de concordancia y confiabilidad. Los valores del coeficiente mean absolute error correspondiente a 0.013 y el valor de root mean squared error 0.113 evidencian el grado óptimo de exactitud de la predicción, por consiguiente, se considera que el modelo propuesto predice de manera eficiente la deserción estudiantil universitaria.

CAPÍTULO IX: CONCLUSIONES, LIMITACIONES Y TRABAJOS FUTUROS

9.1 Conclusiones

En esta etapa se presenta las principales conclusiones establecidas en el presente trabajo de investigación.

- La revisión de la literatura es considerada como la base principal de la investigación, de ella se desprendieron las concepciones empíricas, teóricas y metodológicas de la investigación. Como resultado del proceso de la revisión integral de revisión de literatura se identificó 112 factores que influyen en la deserción de los estudiantes en las universidades, 10 técnicas utilizadas para el preprocesamiento de los datos, 10 técnicas para la selección de factores y 14 técnicas de minería de datos que fueron utilizadas para predecir la deserción en las universidades.
- Se identificaron 11 nuevos factores de deserción basados en el estudio de las teorías educacionales y administrativas. Se demostró experimentalmente que estos factores identificados tienen un alto nivel de significancia estadística y determinan un alto grado de causalidad con la deserción en las universidades. Cuando todos los factores identificados se activan, la probabilidad de que un estudiante abandone la universidad corresponde al 95 %.
- Con el auge de las tecnologías que hoy en día se encuentran al alcance de todo el mundo, especialmente de la comunidad universitaria, se pudo establecer factores tecnológicos que influyen negativamente en la decisión que toman los estudiantes de abandonar sus estudios

universitarios. Se identificó 7 factores tecnológicos que sirvieron como variables de ingreso a los modelos de predicción: árbol de decisión y regresión logística. Los resultados arrojaron que el método de arboles de decisión presenta la mayor tasa de precisión de la predicción del 91.70 %.

- El uso de métodos de minería de datos, especialmente la aplicación de árbol de decisión basados en datos de los estudiantes universitarios relacionados con las adicciones, permitieron determinar que estos factores pueden ser considerados como predictores de la deserción. Los resultados del proceso experimental muestran un porcentaje de predicción del 97.95 % a través del algoritmo CHAID.
- La aplicación de redes neuronales para predecir la deserción estudiantil universitaria mediante el uso de los métodos perceptrón multicapa y función de base radial presentaron como resultado una tasa de precisión del 96.3 % y del 96.8 % respectivamente. Se puede evidenciar que estos métodos de minería de datos son fiables para la identificación temprana de estudiantes con riesgo de desertar de las universidades.
- Para mejorar la tasa de precisión de la predicción de la deserción estudiantil universitaria, en su primera fase del proceso experimental, se aplicó técnicas de minería de datos que dieron como resultado tasas de precisión óptimas que sobrepasan el 90 %, Sin embargo, el diseño de un método híbrido construido través del uso de mata clasificadores de minería de datos permitió mejorar la tasa de precisión al 99.5 %, con data set compuesto de 6000 estudiantes universitarios, lo que indica que el modelo de predicción de la deserción estudiantil universitaria propuesto es óptimo en términos de concordancia y confiabilidad.

9.2 Limitaciones

Se tallan las limitaciones encontradas en el desarrollo de la propuesta de investigación, que van desde el problema, aporte y la validación de resultados.

- La literatura permitió determinar que no existe una definición clara sobre el concepto de deserción y la conceptualización existente que indique cuando un estudiante debe ser considerado como desertor.
- El estado del arte fue limitado a realizarse basado en revistas y conferencias con factor de impacto en inglés y en un periodo de tiempo de los 12 años.
- Los métodos de minería de datos y el meta clasificador propuesto se validaron en universidades públicas del Ecuador.

9.3 Trabajos Futuros

- Determinar factores de deserción desde la perspectiva de los expertos. Es importante conocer la percepción que tiene los especialistas en temas de deserción en las universidades, con la finalidad de establecer nuevos factores que influyan negativamente en la decisión que toman los estudiantes de abandonar sus estudios.
- La aplicación de nuevas técnicas de machine learning como el Deep Learning podrían ser consideradas como una alternativa de mejora que permita la comparación de los algoritmos tradicionales de minería de datos.
- La propuesta de una metodología de predicción de la deserción estudiantil universitaria orientada a estudiantes universitarios con necesidades especiales.
- Plantear estrategias para mitigar los efectos negativos de la deserción y potenciar la permanencia estudiantil en las universidades.

9.4 Publicaciones:

Se detalla las publicaciones realizadas en el proceso de formación doctoral.

9.4.1 Conferencias:

- Factors to Predict at the Universities: A case of study in Ecuador, IEEE. In Global Engineering Education Conference, 2018. Spain, pp 1238-1242.

9.4.1 Revistas

- Prediction of University Dropout through Technological Factors. Revista Espacios, pp 39, vol 52.
- Factors that Influence Undergraduate University Desertion According to Students Perspective. International Journal of Engineering and Technologies, 2018, pp. 1585-1602.
- Neural Networks to Predict at the Universities. International Journal of Machine Learning and Computation, 2018. pp. 149-143.
- Decisión Trees for the early Identification of University Students of Risk of Desertion. International Journal of Engineering and Technologies, 2018. Pp. 51-54.
- Predicting University Dropout trough Data Mining: A Systematic Literature. Indian Journal of Science and Technology.2018. pp. 1585-1602, vol 10.
- Data Mining Approach to Predict University Student Dropout. Medwell Journals, 2018. Aceptado para publicación.

REFERENCIAS BIBLIOGRAFICAS:

- Abu-Oda, G. S., & El-Halees, A. M. (2015). Data mining in higher education: university student dropout case study. *Data mining in higher education: university student dropout case study*, 5(1).
- Al.Barrak, M.A., & Alrazgan, M. S. (2015). Predicting Students' Performance through Classification: A Case Study. *Journal of Theoretical & Applied Information Technology*, 75(2).
- Alarcon, G. M., & Edwards, J. M. (2013). Ability and motivation: Assessing individual factors that contribute to university retention. *Journal of Educational Psychology*, 105(1), 129.
- Alkan, N (2014). Humor, loneliness and acceptance: Predictors of University drop-out intentions. *Procedia. Social and Behavioral Sciences*, 152, 1079-1086.
- Al-Omari, A. I., & Bouza, C. N. (2014). Review of ranked set sampling: modifications and applications. *Investigación Operacional*, 35(3), 215-235.
- Arbid, D.C., & Meiran, N. (2015). Performance on the antisaccade task predicts dropout from cognitive training. *Intelligence*, 49, 25-31.
- Archambault, M., Janosz, V., Dupere, M., Mrault, C., & Andrew, M. (Individual, social and family factors associated with high school dropout among low SES youth: Differential effects as a function of immigrant status. *British Journal of Education Psychology*.
- Argote, I., & Jimenez, R. (2014). Detección de patrones de deserción en los programas de pregrado de la universidad mariana de san juan de pasto, aplicando el proceso de descubrimiento de conocimiento sobre base de datos (kdd) y su implementación en modelos matemáticos de predicción. In *Congresos CLABES*.
- Arifin, M.H. (2016). Exploring factors in contributing student progress in the Open University. *International Journal of Information and Education Technology*, 6(1), 29.

- Arulampalam, W., Naylor, R.A., & Smith, J.P. (2005). Effects of in-class variation and student rank on the probability of withdrawal: cross-section and time-series analysis for UK university student. *Economics of Education Review*, 24(3), 251-262.
- Arulampalam, W., Naylor, R.A., & Smith, J.P. (2007). Dropping out of medical school in the UK: explaining the changes over ten years. *Medical Education*, 41(4), 385-394.
- Astin, A.W. (1975). Preventing students from dropping out. San Francisco: Jossey-Bass.
- Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *arXiv preprint arXiv: 1606.06364*.
- Azis, A.A., Idris, W.M.R.W., Hassan, H., Jusoh, J.A., & Emran, N.A (2018). Implementing Aproiri Algorithm for Predicting Result Analysis. *GSTF Journal of Computing (JoC)*, 2(4).
- Badr. G., Algobail, A., Almutairi, H., & Almutery, M. (2016). Predicting student's performance in university courses: a case study and tool in KSU mathematics department. *Procedia Computer Science*, 82,80-89.
- Ballesteros, H.F.V., Iñiguez, E.G., & Velasco, S.R.M. (2018). Minería de Datos. *RECIMUNDO*, 2(1(Esp)), 339-349.
- Bayer, J., Bydzovská, H., Géryk, J., Obsivac, T., & Popelinsky, L. (2012). Predicting Drop.Out from Social Behaviour of Student of Students. *International Education data Mining Society*.
- Belo, P., & Oliveira, C. (2015). The Relations between Experiences and Expectations with University Dropout. *Procedia-Social and Behavioral Science*, 87, 98-101.
- Black, A. Y., Fleming, N. A., & Rome, E. S. (2012). Pregnancy in adolescents. *Adolescent medicine: State of the art reviews*, 23(1), 123-38.
- Blau, I. (2011). Application use, online relationship types, self-disclosure, and Internet abuse among children and youth: Implications for education

- and Internet safety programs. *Journal of Educational Computing Research*, 45(1), 95-116.
- Brookover, W.B., & Schneider, J.M (1975). Academic environments and elementary school achievement. *Journal of research & Development in Education*.
- Bustos Andreu, H., & Nussbaum, M. (2009). An experimental study of the inclusion of technology in higher education. *Computer Applications in Engineering Education*, 17(1), 100-107.
- Canales, A., & De los Ríos, D. (2007). Factores explicativos de la deserción universitaria. *Calidad en la Educación*, (26), 173-201.
- Canary, J. D., Blizzard, L., Barry, R. P., Hosmer, D. W., & Quinn, S. J. (2017). A comparison of the Hosmer–Lemeshow, Pigeon–Heyse, and Tsiatis goodness-of-fit tests for binary logistic regression under two grouping methods. *Communications in Statistics-Simulation and Computation*, 46(3), 1871-1894.
- Carbonell, X., Chamorro, A., Griffiths, M., Oberst, U., Cladellas, R., & Talarn, A. (2012). Problematic Internet and cell phone use in Spanish teenagers and young students. *Anales de psicología*, 28(3), 789-796.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2008). Análisis de los factores asociados a la deserción estudiantil en la Educación Superior: un estudio de caso Analysis of the Factors Associated with the Drop-out Rate of Students in Higher Education: a Case Study. *Revista de Educación*, 345, 255-280.
- Chen, R. (2012). Institutional characteristics and college student dropout risk: A multilevel event history analysis. *Research in Higher Education*, 53(5), 487-505.
- Chen, R., & DesJardins, S.L. (2008). Exploring the effects of financialaid on the gap in student dropout risk by income level. *Research in higher education*, 49(1), 1-18.
- Ćukušić, M., Garača, Ž., & Jadrić, M. (2014). Online self-assessment and students' success in higher education institutions. *Computers & Education*, 72, 100-109.

- Del Carmen Ibarra, M., & Michalus, J., (2010). Analysis of academic performance using a Logit model. *Rvista de Ingeniería Industrial*, 10.
- Delen, D., Topuz, K., & Eryarsoy, E. (2019). Development of a Bayesian Belief Network-based DSS for Predicting and Understanding Freshmen Student Attrition. *European Journal of Operational Research*.
- Dewan, M. A. A., Lin, F., & Wen, D. (2015, August). Predicting Dropout-Prone Students in E-Learning Education System. In *Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015 IEEE 12th Intl Conf on* (pp. 1735-1740). IEEE.
- Di Domenico, S. I., & Ryan, R. M. (2017). Commentary: Primary emotional systems and personality: An evolutionary perspective. *Frontiers in psychology*, 8, 1414.
- Di Pietro, G., & Cutillo, A. (2008). Degree flexibility and university dropout: The Italian experience. *Economics of Education Review*, 27(5), 546-555.
- Díaz Peralta, C. (2008). Modelo conceptual para la deserción estudiantil universitaria Chilena. *Estudios pedagógicos (Valdivia)*, 34(2), 65-86.
- Díaz, Á. H. F. (2009). Análisis sobre la deserción en la educación superior a distancia y virtual: el caso de la UNAD-Colombia. *Revista de Investigaciones UNAD*, 8(2), 117-149.
- Diccionario de Informática, España, 2014.
- Djulovic, A., & Li, D. (2013). Towards freshman retention prediction: a comparative study. *International Journal of Information and Education Technology*, 3(5), 494.
- Duarte, R., Ramos-Pires, A., & Gonçalves, H. (2014). Identifying at-risk students in higher education. *Total quality management & business excellence*, 25(7-8), 944-952.
- Dueñas, R. A., & Valdivia, M. J. R. (2016). Bullying y poder. *Investigación Valdizana*, 10(3), 119-122.
- Duque, L. C. (2014). A framework for analysing higher education performance: students' satisfaction, perceived learning outcomes, and

- dropout intentions. *Total Quality Management & Business Excellence*, 25(1-2), 1-21.
- Eckert, K. B., & Suénaga, R. (2015). Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. *Formación universitaria*, 8(5), 03-12.
- Elffers, L. (2013). Staying on track: behavioral engagement of at-risk and non-at-risk students in post-secondary vocational education. *European Journal of Psychology of Education*, 28(2), 545-562.
- Elias-Andreu, M. (2008). Los abandonos universitarios: retos ante el Espacio Europeo de Educación Superior.
- Fei, M., & Yeung, D. Y. (2015, November). Temporal models for predicting student dropout in massive open online courses. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on* (pp. 256-263). IEEE.
- Fiegehen, L. (2005). Repitencia y deserción universitaria en América Latina. *presentado en el Seminario de Educación Superior de América Latina y el Caribe. Capítulo, 11.*
- Garcia, M., Sémpertegui, E. (2018). The labor insertion in Higher Education. *The Latin American Perspective Education*, 21
- Garrido, B. Student desertion in the master in University Teaching for first – year student. *Análisis de la Realidad Nacional*, 102(44), 44.
- Gartner Isaza, L., Dussán Lubert, C., & Montoya, D. M. (2016). Characterization of Student Dropout at the University of Caldas in the 2009-2013 Period. Analysis from the System for the Prevention of the Dropout of Higher Education. SPADIES. *Revista Latinoamericana de Estudios Educativos (Colombia)*, 12(1).
- Gencer, S. L., & Koc, M. (2012). Internet abuse among teenagers and its relations to internet usage patterns and demographics. *Journal of Educational Technology & Society*, 15(2), 25-36.
- Ghamari, F., Mohammadbeigi, A., Mohammadsalehi, N., & Hashiani, A. A. (2011). Internet addiction and modeling its risk factors in medical students, Iran. *Indian journal of psychological medicine*, 33(2), 158.

- Giannoulis, C. e Ishizaka, A. (2010). Un sistema de soporte de decisiones basado en la web con ELECTRE III para un ranking personalizado de universidades británicas. *Sistemas de soporte de decisiones*, 48 (3), 488-497.
- Girón, F. (2014). Factores de Riesgo que Ocasionaron la Deserción de Estudiantes de la Facultad de Ingeniería de la Universidad Rafael Ladívar.
- Gitto, L., Minervini, L. F., & Monaco, L. (2016). University dropouts in Italy: Are supply side characteristics part of the problem?. *Economic Analysis and Policy*, 49, 108-116.
- González-Flores, M., Heracleous, M., & Winters, P. (2012). Leaving the safety net: an analysis of dropouts in an urban conditional cash transfer program. *World Development*, 40(12), 2505-2521.
- Guarín, C. E. L., Guzmán, E. L., & González, F. A. (2015). A model to predict low academic performance at a specific enrollment using data mining. *IEEE Revista Iberoamericana de tecnologías del Aprendizaje*, 10(3), 119-125.
- Gujarati, D., & Porter, D. (2010). *Econometric* (quinta edición).
- Gutierrez, M., Ulloa, H., Simancas, I., Velasco, I. (2017). Causas de la Deserción Escolar en los alumnos de primer año de la licenciatura de contaduría en la modalidad de semiescolarizado de la unidad académica de contaduría y administración.
- Hallinger, P., Hosseingholizadeh, R., Hashemi, N., & Kouhsari, M. (2018). Do beliefs make a difference? Exploring how principal self-efficacy and instructional leadership impact teacher efficacy and commitment in Iran. *Educational Management Administration & Leadership*, 46(5), 800-819.
- Henríquez, C., & Escobar, R. (2016). Construcción de un modelo de alerta temprana para la detección de estudiantes en riesgo de deserción de la Universidad Metropolitana de Ciencias de la Educación. *Revista mexicana de investigación educativa*, 21(71), 1221-1248.
- Heredia, D., Amaya, Y., & Barrientos, E. (2015). Student dropout predictive model using data mining techniques. *IEEE Latin America Transactions*, 13(9), 3127-3134.

- Hershkovitz, A., & Nachmias, R. (2011). Online persistence in higher education web-supported courses. *The Internet and Higher Education, 14*(2), 98-106.
- Herzog, S. (2005). Measuring determinants of student return vs. dropout/stopout vs. transfer: A first-to-second year analysis of new freshmen. *Research in higher education, 46*(8), 883-928.
- Himmel, E. (2002). Modelo de análisis de la deserción estudiantil en la educación superior. *Calidad en la Educación, (17)*, 91-108.
- Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems, 101*, 1-11.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Hovdhaugen, E. (2009). Transfer and dropout: Different forms of student departure in Norway. *Studies in Higher Education, 34*(1), 1-17.
- Hoyt, C. R. (1986). The Impact of the Tax Reform Act of 1986 on Legal Education and Law Faculty. *J. Legal Educ., 36*, 568.
- Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469-478.
- Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education, 61*, 133-145.
- Human-Vogel, S., & Rabe, P. (2015). Measuring self-differentiation and academic commitment in University students: A case study of education and engineering students. *South African Journal of Psychology, 45*(1), 60-70.
- Iepsen, E. F., Bercht, M., & Reategui, E. (2013, October). Detection and assistance to students who show frustration in learning of algorithms. In *Frontiers in Education Conference, 2013 IEEE* (pp. 1183-1189). IEEE.
- Ito, T. (2009). Caste discrimination and transaction costs in the labor market: Evidence from rural North India. *Journal of development Economics, 88*(2), 292-300.

- Jadrić, M., Garača, Ž., & Čukušić, M. (2010). Student dropout analysis with application of data mining methods. *Management: journal of contemporary management issues*, 15(1), 31-46.
- Jantke, K. P., & Drefahl, S. (2016). *Theory of mind modeling and induction: Ausdrucksfähigkeit und Reichweite*. ADISY Technical Report 03/2016. Weimar, Germany: ADISY Consulting GmbH.
- Jia, P., & Maloney, T. (2015). Using predictive modelling to identify students at risk of poor university outcomes. *Higher Education*, 70(1), 127-149.
- Kaufman, P., Alt, M. N., & Chapman, C. D. (2004). Dropout Rates in the United States: 2001. Statistical Analysis Report NCES 2005-046. *US Department of Education*.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), 1-26.
- Kittinger, R., Correia, C. J., & Irons, J. G. (2012). Relationship between Facebook use and problematic Internet use among college students. *Cyberpsychology, Behavior, and Social Networking*, 15(6), 324-327.
- Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 60-65).
- Koonce, D. A., & Hening, D. A. (2009). Data Imputation to Identify Potential Dropouts. In *IIE Annual Conference. Proceedings* (p. 246). Institute of Industrial and Systems Engineers (IISE).
- Krstić, K., Lazarević, L. B., & Ilić, I. S. (2016). Dropout as a result of education with no space for diversity. In *Open Spaces for Interactions and Learning Diversities* (pp. 129-136). SensePublishers, Rotterdam.
- Lam-On, N., & Boongoen, T. (2014, December). Using cluster ensemble to improve classification of student dropout in Thai university. In *Soft Computing and Intelligent Systems (SCIS), 2014 Joint 7th International Conference on and Advanced Intelligent Systems (ISIS), 15th International Symposium on* (pp. 452-457). IEEE.

- Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area—Application to student data from Open University of Madrid, UDIMA. *Computers & Education*, 72, 23-36.
- Larsen, MR, Sommersel, HB y Larsen, MS (2013). *Evidencia sobre fenómenos de abandono en las universidades*.
- Leal, A., Torrealba, A. (2017). Motivational actors an integrative approach in the academic performance of the subject office asministration in the mention: education for work sub-commercial area of the Faculty of Education Sciences at the University of Carabobo.
- Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & education*, 48(2), 185-204.
- Li, W., Gao, M., Li, H., Xiong, Q., Wen, J., & Wu, Z. (2016, July). Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In *Neural Networks (IJCNN), 2016 International Joint Conference on* (pp. 3130-3137). IEEE.
- Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016, April). Big data application in education: dropout prediction in edx MOOCs. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)* (pp. 440-443). IEEE.
- Lin, S. H. (2012). Data mining for student retention management. *Journal of Computing Sciences in Colleges*, 27(4), 92-99.
- Lopez, V. (2004). Strategies to overcome university desertion. *Educación y Educadores*, vol 7, pp177.
- Lye, C. T., Ng, L. N., Hassan, M. D., Goh, W. W., Law, C. Y., & Ismail, N. (2010). Predicting Pre-university student's Mathematics achievement. *Procedia-Social and Behavioral Sciences*, 8, 299-306.
- Lykourantzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, 53(3), 950-965.

- Magadi, M., Otieno, A. T. A., Obare, F., & Taffa, N. (2004). Comparing maternal health indicators between teenagers and older women in sub-Saharan Africa.
- Márquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 8(1), 7-14.
- Martin, G., Martinez, R., Nieto, F & Nuñez, V,. Acercamiento de las Teorías del Aprendizaje en la Educación Superior. UNIANDES EPISTEME, vol 4, pp.48-60, 2017.
- Martinho, V. R., Nunes, C., & Minussi, C. R. (2013, September). Prediction of school dropout risk group using neural network. In *Computer science and information systems (FedCSIS), 2013 federated conference on* (pp. 111-114). IEEE.
- Melguizo, T., Sanchez, F., & Velasco, T. (2016). Credit for low-income students and access to and academic performance in higher education in Colombia: A regression discontinuity approach. *World Development*, 80, 61-77.
- Méndez-Estrada, V. H., & Llosa, Z. B. (2012). Uso de tecnologías de la informática y comunicación (TIC) para disminuir la deserción de egresados en posgrados semipresenciales. *UNED Research Journal*, 4(2), 231-238.
- Miranda, A. B. L. (2017). Academic excellence indicators, the perspective of postgraduate students. *Journal for Educators, Teachers and Trainers*, 8(1).
- Molina Vargas, J. A. S. O. N., Torres Pinzón , C., & Restrepo Patiño, C. (2008). Técnicas de Inteligencia Artificial para la Solución de Laberintos de Estructura Desconocida. *Scientia et Technica*, 14(39).
- Montaño Moreno, J. J. (2017). Redes neuronales artificiales aplicadas al análisis de datos.
- Moreno, C., & Molina, Y. (2014). Evaluación del proceso de retención: desde los que enseñan y aprenden en una educación mediada por las TIC. In *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación, Buenos Aires, Argentina*.

- Mueller, T., Tevendale, H. D., Fuller, T. R., House, L. D., Romero, L. M., Brittain, A., & Varanasi, B. (2017). Teen pregnancy prevention: Implementation of a multicomponent, community-wide approach. *Journal of Adolescent Health, 60*(3), S9-S17.
- Murillo, J. (2006). Cuestionarios y escalas de actitudes. *Madrid: Universidad Autónoma de Madrid, 2-16.*
- Natek, S., & Zwilling, M. (2014). Student data mining solution–knowledge management system related to higher education institutions. *Expert systems with applications, 41*(14), 6400-6407.
- Navarro, J., (2014).The Labor Insertion in Higher Education. The Latin American Perspective Universitat Autònoma de Barcelona.
- Navia, C. H., & Tamayo, O. E. (2018). 2B019 Modelos mentales sobre ambiente en estudiantes indígenas de educación básica. *Tecné Episteme y Didaxis TED.*
- Nistor, N., & Neubauer, K. (2010). From participation to dropout: Quantitative participation patterns in online university courses. *Computers & Education, 55*(2), 663-672.
- Oeda, S., & Hashimoto, G. (2017). Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes. *Procedia Computer Science, 112*, 614-621.
- O'Neill, L. D., Christensen, M. K., Vonsild, M. C., & Wallstedt, B. (2014). Program specific admission testing and dropout for sports science students: a prospective cohort study. *Dansk Universitetspædagogisk Tidsskrift, 9*(17), 55-70.
- Organización de las naciones Unidad para la Educación, la Ciencia y la Cultura. Educación 2030, Educadores, 2013.
- Organización para la Cooperación y el Desarrollo Económico (OECD), 2008.
- Organización para la Cooperación y el Desarrollo Económico (OECD), 2014.
- Ortiz-Lozano, J. M., Rua-Vieites, A., Bilbao-Calabuig, P., & Casadesús-Fa, M. (2018). University student retention: Best time and data to identify undergraduate students at risk of dropout. *Innovations in Education and Teaching International, 1-12.*

- Oseguera, L., & Rhee, B. S. (2009). The influence of institutional retention climates on student persistence to degree completion: A multilevel approach. *Research in Higher Education*, 50(6), 546-569.
- Pal, S. (2012). Mining educational data to reduce dropout rates of engineering students. *International Journal of Information Engineering and Electronic Business*, 4(2), 1.
- Panoráma de la Educación, 2013
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4).
- Páramo, G., & Correa, C. (1999). Deserción estudiantil universitaria. Conceptualización. *Revista Universidad EAFIT*, 114, 65-78.
- Park, J. H., & Choi, H. J. (2009). Factors influencing adult learners' decision to drop out or persist in online learning. *Journal of Educational Technology & Society*, 12(4).
- Paura, L., & Arhipova, I. (2014). Cause analysis of students' dropout rate in higher education study program. *Procedia-Social and Behavioral Sciences*, 109, 1282-1286.
- Pereira, R. T., Romero, A. C., & Toledo, J. J. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Revista vínculos*, 10(1), 373-383.
- Pérez López, C. (2004). Técnicas de análisis multivariante de datos. *Aplicaciones con SPSS, Madrid, Universidad Complutense de Madrid*.
- Pérez, A. M., Escobar, C. R., Toledo, M. R., Gutierrez, L. B., & Reyes, G. M. (2018). Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O' Higgins. *Educação e Pesquisa*, 44, 1-23.
- Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: A case for the inclusion of student engagement variables. *Remedial and Special Education*, 27(5), 276-292.
- Rodríguez Vignoli, J. (2005). Reproducción en la adolescencia: el caso de Chile y sus implicaciones de política. *Revista de la CEPAL*.

- Rodríguez, A. A., & Vindas, M. A. S. (2005). La deserción estudiantil en la educación superior: el caso de la Universidad de Costa Rica. *Actualidades investigativas en educación*, 5.
- Rodríguez-Gómez, D., Feixas, M., Gairín, J., & Muñoz, J. L. (2012). Understanding Catalan University dropout from a comparative approach. *Procedia-Social and Behavioral Sciences*, 46, 1424-1429.
- Rojas, L. (2013). Comprehensive behavior and school bullying in high school student.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2010). *Handbook of educational data mining*. CRC press.
- Roso-Bas, F., Jimenez, A. P., & García-Buades, E. (2016). Emotional variables, dropout and academic performance in Spanish nursing students. *Nurse education today*, 37, 53-58.
- Sanchez, G., Navarro, W., and Garcia, A. (2002). Factores de deserción estudiantil en la Universidad Surcolombiana.
- Sangodiah, A., Beleya, P., Muniandy, M., Henhg, L. E., & Ramedran, C. H. A. R. L. E. S. (2015). Minimizing Student Attrition In Higher Learning Institutions in Malaysia Using Support Vector Machine. *Journal of Theoretical & Applied Information Technology*, 71(3).
- Saranya, A., & Rajeswari, J. (2016). Enhanced Prediction of Student Dropouts Using Fuzzy Inference System And Logistic Regression. *ICTACT Journal on Soft Computing*, 6(2).
- Simpson, O. (2013). Student retention in distance education: are we failing our students?. *Open Learning: The Journal of Open, Distance and e-Learning*, 28(2), 105-119.
- Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive Modeling of Student Dropout Indicators in Educational Data Mining using

- Improved Decision Tree. *Indian Journal of Science and Technology*, 9 (4).
- Sneyers, E., & De Witte, K. (2017). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. Evidence from the Netherlands. *Studies in Higher Education*, 42(2), 354-389.
- Stewart, S., Lim, D. H., & Kim, J. (2015). Factors influencing college persistence for first-time students. *Journal of Developmental Education*, 12-20.
- Stratton, L. S., O'Toole, D. M., & Wetzel, J. N. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of education review*, 27(3), 319-331.
- Suhirman, S., Herawan, T., Chiroma, H., & Zain, J. M. (2014). Data mining for education decision support: a review. *International Journal of Emerging Technologies in Learning (iJET)*, 9(6), 4-19.
- Sultana, S., Khan, S., & Abbas, M. A. (2017). Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. *International Journal of Electrical Engineering Education*, 54(2), 105-118.
- Tan, M., & Shao, P. (2015). Prediction of student dropout in e-Learning program through the use of machine learning method. *International Journal of Emerging Technologies in Learning (iJET)*, 10(1), 11-17.
- Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2), 321-330.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1), 89-125.
- Torres, C. Z., Ramos, C. A., & Moraga, J. L. (2016). Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos. *Ciencia Amazónica:(Iquitos)*, 6(1), 73-84.
- Tumen, S., Shulruf, B., & Hattie, J. (2008). Student pathways at the university: Patterns and predictors of completion. *Studies in Higher Education*, 33(3), 233-252.

- Vargas, J. M., Pinzón, C. T., & Patiño, C. R. (2008). Técnicas de inteligencia artificial para la solución de laberintos de estructura desconocida. *Scientia et technica*, 2(39).
- Vélez, E. C., Gómez, S. G., Portilla, K. G., & Velásquez, J. V. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de economía*, (65), 9-36.
- Vogel, C., Hochberg, J., Hackstein, S., Bockshecker, A., Bastiaens, T. J., & Baumöl, U. (2018, June). Dropout in Distance Education and how to Prevent it. In *EdMedia+ Innovate Learning*(pp. 1788-1799). Association for the Advancement of Computing in Education (AACE).
- Vries, W. D., León Arenas, P., Romero Muñoz, J. F., & Hernández Saldaña, I. (2011). ¿Desertores o decepcionados? Distintas causas para abandonar los estudios universitarios. *Revista de la educación superior*, 40(160), 29-49.
- Willcoxson, L., Cotter, J., & Joy, S. (2011). Beyond the first-year experience: the impact on attrition of student experiences throughout undergraduate degree studies in six diverse universities. *Studies in Higher Education*, 36(3), 331-352.
- Willging, P. A., & Johnson, S. D. (2009). Factors that influence students' decision to dropout of online courses. *Journal of Asynchronous Learning Networks*, 13(3), 115-127.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wolf, P. S., David, A., Butler-Barnes, S. T., & Zile-Tamsen, V. (2017). American Indian/Alaskan Native college dropout: Recommendations for increasing retention and graduation. *Journal on Race, Inequality, and Social Mobility in America*, 1(1), 1.
- Wray, J., Barrett, D., Aspland, J., & Gardiner, E. (2012). Staying the course: Factors influencing pre-registration nursing student progression into Year 2—A retrospective cohort study. *International Journal of Nursing Studies*, 49(11), 1432-1442.
- Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit

- through stacking generalization. *Computers in Human Behavior*, 58, 119-129.
- Yasmin, D. (2013). Application of the classification tree model in predicting learner dropout behaviour in open and distance learning. *Distance Education*, 34(2), 218-231.
- Ye, C., & Biswas, G. (2014). Early prediction of student dropout and performance in MOOCs using higher granularity temporal information. *Journal of Learning Analytics*, 1(3), 169-172.
- Yi, H., Zhang, L., Yao, Y., Wang, A., Ma, Y., Shi, Y., ... & Rozelle, S. (2015). Exploring the dropout rates and causes of dropout in upper-secondary technical and vocational education and training (TVET) schools in China. *International Journal of Educational Development*, 42, 115-123.
- Yoguez, A. (2009). How world-class universities are evaluated. *Revista de Educación Superior*, 38, 113-120.
- Yu, W. H., & Su, K. H. (2006). Gender, sibship structure, and educational inequality in Taiwan: Son preference revisited. *Journal of Marriage and Family*, 68(4), 1057-1068.
- Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and E-learning*, 17(1), 118-133.
- Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, 27, 44-53.

ANEXOS A. Encuesta sobre factores que inciden en la deserción estudiantil universitaria

Objetivo: Conocer información relacionada con factores que influyen en la decisión de los estudiantes de abandonar las aulas universitarias, con el propósito de presentar a los administradores de las instituciones de educación superior información valiosa que permita la implementación de estrategias para la reducción de las tasas de deserción y culminación de los estudios universitarios con éxito.

Instrucciones: Siendo la deserción estudiantil universitaria el abandono prematuro de un programa de estudios antes de alcanzar un título o grado, se considera un tiempo suficientemente largo como para descartar la posibilidad de que el estudiante se reincorpore a las aulas de clase. Esta encuesta permitirá descubrir los Factores que afectan a la Deserción Estudiantil Universitaria en el Primer Año. Está dirigida para los estudiantes de pregrado de universidad, las respuestas que se obtengan de este cuestionario serán utilizadas estrictamente con fines de investigación. La encuesta está dividida en 3 secciones. La sección 1 está relacionada con la caracterización de las instituciones y de los estudiantes que llenan el cuestionario. La sección 2 corresponde a preguntas que determinen la influencia de los factores que inciden en la deserción. La sección 3 contiene preguntas que complementan el estudio.

1. Información general del entrevistado y de la institución

Número	Pregunta
1	Ubique la facultad a la que pertenece: <ul style="list-style-type: none"> • Ciencias Humanas y de Educación • Ciencias Administrativas • Ciencias de la Ingeniería y Aplicadas • Ciencias Agropecuarias
2	Carrera a la que pertenece
3	Género <ul style="list-style-type: none"> • Femenino • Masculino
4	Edad
5	Estado civil <ul style="list-style-type: none"> • Soltero • Casado • Divorciado • Unión libre • Viudo
6	Etnia
7	Nivel de educación del padre
8	Nivel de educación de la madre
9	Tipo de hogar <ul style="list-style-type: none"> • Madre y hermanos • Padre y hermanos • Padre, madre y hermanos • Esposa e hijos • Otros familiares

2.- Factores de deserción

Conteste las preguntas considerando la escala del 1 al 5 el nivel de influencia de cada factor:

1= No influye

2 = Baja influencia

3 = Mediana influencia

4 = Alta influencia

5 = Influye totalmente

Número	Pregunta	Escala Likert
1	¿Cuál es grado de influencia que tiene el factor compromiso del docente con el estudiante con la deserción en las universidades?	1 2 3 4 5
2	¿Qué nivel de influencia considera usted que tienen los siguientes vicios que adquiere el estudiante durante sus años de estudio con la deserción?	1 2 3 4 5
3	¿Considera usted que el limitado conocimiento para la utilización de tecnologías especializadas de la carrera influye negativamente en la decisión de los estudiantes de abandonar la universidad?	1 2 3 4 5
4	¿Considera usted que si un estudiante se siente comprometido económicamente con la familia y es hijo primogénito influiría en la deserción de abandonar la universidad?	1 2 3 4 5
5	¿Considera usted que una baja categorización de la universidad en el ranking universitario influye negativamente en la deserción de los estudiantes en las universidades?	1 2 3 4 5
6	¿Si la universidad en la que usted estudia baja de categoría en el ranking de las universidades, usted optaría por cambiarse de universidad?	1 2 3 4 5
7	¿Considera usted que el embarazo planificado o no planificado influye en la decisión que tienen los estudiantes de abandonar la universidad?	1 2 3 4 5
8	¿Cuál es el nivel de influencia que tiene el número de hijos del estudiante con la decisión de abandonar la universidad?	1 2 3 4 5
9	¿Cuál es el nivel de influencia que tienen los grupos que componen la comunidad universitaria respecto al factor bullying con la deserción estudiantil universitaria?	1 2 3 4 5
10	¿Cuál es el nivel de influencia que tienen las agresiones machistas/feministas con la deserción en las universidades?	1 2 3 4 5
11	¿Cuál es el nivel de influencia que tienen los componentes del factor percepción del estudiante sobre la inserción del campo laboral con la deserción en las universidades?	1 2 3 4 5
12	¿Cuál es el nivel de influencia que tienen los componentes del factor adaptación a las metodologías de formación universitaria con la deserción en las universidades?	1 2 3 4 5

3.- Preguntas que complementan el estudio

Número	Pregunta	Escala
1	¿La carrera que usted está estudiando llena sus expectativas?	<ul style="list-style-type: none">• Sí• No
2	¿El interés por culminar la carrera en la que está estudiando es?	<ul style="list-style-type: none">• Alto• Medio• Bajo
3	¿Ha pensado usted en algún momento en dejar sus estudios universitarios?	<ul style="list-style-type: none">• Sí• No

ANEXOS B. Producción Científica sobre la Investigación

IEEE.org | IEEE Xplore Digital Library | IEEE-SA | IEEE Spectrum | More Sites Cart (0) | Create Account | Personal Sign In

IEEE Xplore® Digital Library Institutional Sign In

Browse ▾ My Settings ▾ Get Help ▾ Subscribe

All ▾ Enter keywords or phrases (Note: Searches metadata only by default. A search for 'smart grid' = 'smart AND grid') Q

[Advanced Search](#) | [Other Search Options](#) ▾

Conferences > 2018 IEEE Global Engineering ... Advertisement

Factors to predict dropout at the universities: A case of study in Ecuador

Publisher: IEEE

2 Author(s) Alban Mayra · David Mauricio [View All Authors](#)

112 Full Text Views

Abstract

The dropout at the universities has become a concern in several countries around the world, its high rates generate negative consequences for students and organizations. Based on the analysis of the educative, organizational theories, and the logic reasoning were established 11 factors that influenced in the dropout. This research as an objective to design a model to determine new factors to predict the dropout in which the dimension of analysis were the students, the institutions, the academic context and the social and economic environment. Additionally, trying through the use of Logistical Regression, Decision Tree and Support Vector Machine if the proposed factors are related and or may contribute predicting the dropout at the universities of Ecuador.

Published in: 2018 IEEE Global Engineering Education Conference (EDUCON)

Date of Conference: 17-20 April 2018 **INSPEC Accession Number:** 17803009

[Authors](#)

Need Full-Text
access to IEEE Xplore for your organization?
[REQUEST A FREE TRIAL >](#)

More Like This

Student pass rates prediction using optimized support vector machine and decision tree
2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)
Published: 2018

Colleges Employment Forecasting by Least Squares Support Vector Machine
2012 International Conference on Computer Science and Electronics Engineering
Published: 2012

[View More](#)

IEEE websites place cookies on your device to give you the best user experience. By using our websites, you agree to the placement of these cookies. To learn more, read our [Privacy Policy](#). [Accept & Close](#)

16:22
23/08/2019

Factors to Predict Dropout at the Universities: A case of study in Ecuador

Alban Mayra

Faculty of Computer Science and Computer Systems
Technical University of Cotopaxi
Latacunga, Ecuador
mayra.alban@utc.edu.ec

David Mauricio

Faculty of Postgraduate of Systems and Informatics
National University San Marcos
Lima, Perú
dmauricios@unmsm.edu.pe

Abstract—The dropout at the universities has become a concern in several countries around the world, its high rates generate negative consequences for students and organizations. Based on the analysis of the educative, organizational theories, and the logic reasoning were established 11 factors that influenced in the dropout. This research as an objective to design a model to determine new factors to predict the dropout in which the dimension of analysis were the students, the institutions, the academic context and the social and economic environment. Additionally, trying through the use of Logistical Regression, Decision Tree and Support Vector Machine if the proposed factors are related and or may contribute predicting the dropout at the universities of Ecuador.

Keywords—university student desertion, factor, data mining

I. INTRODUCTION

The dropout is one of the problems that affect to the majority of the institutions of high education around the world, it turns into a topic that causes controversy in the educative environment in which are involved administrators, professors, and students.

Nowadays, there is high dropout rates in the university education system, this can be evidenced by the data officially presented in the annual reports that Government Agencies worldwide show on the results of academic and administrative management in education [1]. In the India the factor of dropout was 15% in Belgium, the dropout rate was 26, 9% in the United States reached 24%. According to the United Nations Organization in some Latin American countries as Colombia and Ecuador, the dropout overcame 40%, while in Costa Rica the dropout rate reached 50% and in Brazil was 54% approximately.

As a result, university student desertion is related to a swarm of factors that interact with each other [2] and have a positive or negative influence on students' decision to stay or leave the university classroom [3]. Although the empirical evidence on factors that affect dropout is broad, most research has focused on studying the internal or external characteristics of students [4] leaving aside aspects of the academic, social, economic and institutional context in many cases it can affect the university dropout and require more exploration.

obtaining the probability of incidence that has specific factors in the desertion of universities, based on the nature of their variables [5].

Based on the need to establish actions that allow the decrease of student desertion in universities, this article aims to try whether the proposed new factors affect the dropout of university students in Ecuador and verify whether these factors were analyzed in the units personal, academic, social, economic and institutional can predict university dropout. The results of the research will help the administrators of the Ecuadorian universities to promote changes in their academic policies and strategies in order to reduce dropout rates.

This article is organized into five sections. In section two shows the literature review. The method applied to the development of this research is in section three, section four explains the experimental process and the discussion of results, and finally, section five presents the conclusions.

II. LITERATURE REVIEW

The abandonment of students in universities is considered a problem that affects higher education institutions [6]. Currently, school dropout rates are analyzed as quality criteria in evaluation and accreditation processes of institutions of higher education [6]. In many cases, these control processes applied by the Government Agencies to universities imply economic and social changes [7] that they affects the students, institutions, and governments [8].

The review of the literature allows describing the determination of factors that influence the desertion of universities. For Willging & Johnson [9] two theories could provide a complete explanation about factors that affect student desertion, one of which is the theory of the integration of the Tinto model, who studies the motivation of the individual and the ability academic impact in the university dropout.

While the second theory is related to the Bean model, where the behavior of the students are analyzed, based on the beliefs and attitudes of the students. Based on the Theory of Tinto, authors such as Hovdlaugen [10], Elfers [11] & Duque [12] examine dropout factors based on motivation, academic

Predicting University Dropout through Data Mining: A Systematic Literature

Mayra Alban^{1*} and David Mauricio²

¹Technical University of Cotopaxi, Faculty of Computer Science and Computer Systems, Ecuador; mayra.alban@utc.edu.ec

²National University of San Marcos, Artificial Intelligence Group, Perú; dmauricios@unmsm.edu.pe

Abstract

Objectives: To make a systematic review of literature on the prediction of university student dropout through data mining techniques. **Methods/Analysis:** The study was developed as a systematic review of the literature of empirical research results regarding the prediction of university dropout. In this phase, the review protocol, the selection requirements for potential studies and the method for analyzing the content of the selected studies were provided. The classification presented in section 3 allowed answering the main research question. What are the aspects considered in the prediction of university student desertion through data mining? **Findings:** University dropout is a problem which affects universities around the world, with consequences such as reduced enrolment, reduced revenue for the university, and financial losses for the State which funds the studies, and also constitutes a social problem for students, their families, and society in general. Hence the importance of predicting university dropout, that is to say identify dropout students in advance, in order to design strategies to tackle this problem. **Novelty /Improvement:** This is the first work to perform an integral systematic literature review about university dropout prediction through data mining, with studies from 2006–2018.

Keywords: Data Mining, Dropout Factors, Dropout Prediction, Machine Learning, University Student Dropout

1. Introduction

There is currently an increasing interest in researching the topic of university dropout around the world¹, with one of the main concerns being elevated rates of occurrence². Dropout negatively affects institutions in the reduction of enrolment and the non-achievement of institutional objectives³. As a consequence, students, universities and governments are affected in both economic and social terms. Furthermore, dropout becomes a critical topic when university administrators do not possess the tools necessary to identify students who are at risk of leaving the institution. In turn, potential corrective measures are reduced⁴, which might have enabled student retention at higher education institutions⁵. In the same way, the early prediction of student dropout has become a major chal-

sible reason that there are still high university dropout rates may be associated with the fact that most of the prediction models applied to solve this problem are difficult to interpret⁶. A significant effort has been made to close the university dropout gap and thus reduce dropout rates. Nonetheless, this effort has been insufficient⁷; according to the Organization for Economic Cooperation and Development (OECD), in 2016, European dropout rates ranged between 30% and 50%, while in the United States the student dropout rate was 37%⁸.

In some Latin American countries, such as Columbia, dropout rates exceeded 40%, while in Brazil they reached approximately 54%. In Costa Rica, the dropout rate reached 50%⁹, with public universities presenting higher dropout rates than private ones¹⁰. One of the measures to deal with university dropout is based on predicting its



Decision Trees for the Early Identification of University Students at Risk of Desertion

Mayra Albán^{1*}, David Mauricio²

^{1,2}Technical University of Cotopaxi, Faculty of Computer Science, Ecuador

²National University of San Marcos, Artificial Intelligence Group, Perú

*Corresponding author E-mail: mayra.alban@utic.edu.ec

Abstract

The student's dropout at the universities is a topic that has generated controversy in Higher Education Institutions. It has negative effects which cause problems in the social, academic and economic context of the students. One of the alternatives used to predict the dropout at the universities is the implementation of machine learning techniques such as decision trees, known as prediction models that use logical construction diagrams to characterize the behavior of students and identify early students that are at risk of leaving university. Based on a survey of 3162 students, it was possible to obtain 10 variables that have influence into the dropout, that's why, a CHAID decision tree model is proposed that presents the 97.95% of the accuracy in the prediction of the university students' dropout. The proposed prediction model allows the administrators of the universities developing strategies for effective intervention in order to establish actions that allow students finishing their university careers successful.

Keywords: Prediction of college desertion, machine learning, decision trees, CHAID.

1. Introduction

The completion of the university has not always been the norm for the society, in the 1940s, less than half of the US population between the ages of 25 and 29 would have finished the university Ye & Bisway [1]. Although there has existed a concerted effort to close the gap related with the dropout at the universities and decrease its rates, researches that started in 1978 shows that still exists dropout at the universities Abuda & Oda [2] which has caused effects on the economic ambit for Higher Education institutions and governments.

The dropout causes difficulties at the university context [3] and it is considered as an evaluation criterion and an argument of great relevance to incorporate the public policies related with the education at the universities [4]. Although effective systematic changes have been made to solve this problem, students continue to face an educational crisis with the greatest propensity to leave their studies [5]. In addition, dropout generates social consequences in terms of the students' expectations and their families; as well as emotional consequences for the dissonance between the aspirations of young people and their achievements. The important eco-

processes that consider methods of machine learning supervised to discover knowledge.

So, if the above problems are maintained, the high error rates in the accuracy of the prediction will continue. For this reason, it is important to establish a model that allows integrating data, variables and appropriate techniques to accurately predict students at risk of dropping out. In addition, it will allow Higher Education Institutions to have an effective tool to make decisions rightly in relation to the dropout of the universities' students.

The research work is divided into five parts. The literature review is in the second one, the method is developed in the third, the results of the experimental process is considered in the fourth and in the last part are presented the conclusions.

2. Literature Review

The university student desertion is a problem that has been widely studied in the literature. It is possible to demonstrate the efforts made by the researchers around the topic of their prediction to try to mitigate the dropout rates and establish strategies that allow the incorporation of strategies for timely decision making. Several

Factors that Influence Undergraduate University Desertion According to Students Perspective

Mayra Alban^{#1}, David Mauricio^{*2}

[#]Faculty of Engineering in Computer Sciences
Technical University of Cotopaxi, Latacunga, Ecuador
¹mayra.alban@utc.edu.ec

^{*}Faculty in Systems Engineering and Computer Sciences
National University of San Marcos, Lima, Perú
²dmauricos@unmsm.edu.ec

Abstract—University desertion is defined as a withdrawal from the academic system that has a negative impact on the economic and social environment. Desertion is influenced by internal and external factors that affect students, institutions, and governments. A literature review shows empirical evidence concerning research conducted to solve the desertion problem. However, most of the studies we analyzed do not consider students' perceptions, which are a key factor in college dropout rates. For this reason, we propose some new factors that influence university desertion. These factors are supported by a study of 65 organizational and educational theories. To examine the influence of these factors, a case study was conducted with 3773 students at a public university in Ecuador. Logistic regression was used to determine the causal relationship between the identified factors and the students' university desertion. As a result, we established that when all the identified factors are present at the same time, the probability of a student deciding to drop out of college is 95% your paper to be published in the conference proceedings, you must use this document as both an instruction set and as a template into which you can type your own text. If your paper does not conform to the required format, you will be asked to fix it.

Keyword—Educational data mining, logistic regression, university desertion, desertion factors

I. INTRODUCTION

University desertion is a problem that affects most higher education institutions around the world. This topic generates controversy among administrators, professors, and students [1]. Today, high dropout rates are considered a possible deficiency of the undergraduate education system [2].

This can be seen from the academic and administrative indicators presented by management personnel at universities and higher education control agencies. In 2016, the dropout rate in India was 15.9% [3], while in Belgium, it was 26.9%; in the United States, the desertion rate for first and second year students was 44.8% [4], and in Costa Rica, it reached up to 49%. According to the United Nations (UN), the 2016 university desertion rate was 40% in several Latin-American countries, such as Colombia and Ecuador, and was approximately 54% in Brazil.

University desertion is influenced by a set of interacting factors that have a negative impact on students' decisions to drop out of college [5]. A literature review allows us to identify several studies related to factors that influence desertion. Nevertheless, these studies do not consider students' perspectives, which play a key role in university desertion. Consequently, the following study is based on organizational and educational theories and proposes 11 new factors that affect university desertion while taking students' perspectives into consideration. To examine the influence of these factors, an empirical study is performed with 3773 students.

This article is organized into six sections. In section two, the literature review is presented. The methodology applied in this research is described in section three. Sections four and five detail the experimental process and discuss the results, respectively. Finally, in section six, we state our conclusions.

II. LITERATURE REVIEW

University desertion can be defined as the cessation of an academic process and the non-completion of a formative process in higher education [6]. It is a problem that worries governments and has become a weakness in university education due to its high rates, which have negative effects on students and on the economic growth of societies [7]. Specifically, a university deserter is a student who has not exhibited academic activity for two



Prediction of university dropout through technological factors: a case study in Ecuador

Predicción de la deserción estudiantil universitaria a través de factores tecnológicos: un caso de estudio en Ecuador

Mayra Susana ALBAN Talpe 1; David MAURICIO Sánchez 2

Received: 27/06/2018 • Approved: 10/09/2018 • Published 28/12/2018

Contents

[1. Introduction](#)

[2. Methodology](#)

[3. Results](#)

[4. Conclusions](#)

[Bibliographic references](#)

ABSTRACT:

Predicting dropout in universities has become a concern in several countries around the world. With the introduction of new information and communication technologies, new factors have appeared that influence student dropout in universities. This article proposes an approach to machine learning based on logistic regression techniques and decision trees and factors such as Internet addiction, addiction to social networks and addiction to technology, that affect the desertion of students in universities. As a result, it was obtained that the technique with the highest percentage of dropout precision was decision trees with 91.70%.

Keywords: Prediction of dropout, technological factors, machine learning

RESUMEN:

Predicir la deserción en las universidades se ha convertido en una preocupación en varios países del mundo. Con la introducción de las nuevas tecnologías de información y comunicación han aparecido nuevos factores que influyen en la deserción estudiantil en las universidades. Este artículo propone un enfoque de aprendizaje automático basado en las técnicas de regresión logística y árboles de decisión y en los factores adicción al internet, adicción a las redes sociales y adicción a la tecnología que afectan a la deserción de los estudiantes en las universidades. Como resultado se obtiene que la técnica que presenta mayor porcentaje de precisión de la deserción fue árboles de decisión con un 91.70%.

Palabras clave: Predicción de la deserción, factores tecnológicos, aprendizaje automático

1. Introduction

Neural Networks to Predict Dropout at the Universities

Mayra Alban and David Mauricio

Abstract—The university student's dropout is a problem that affects the governments, institutions and students. It has negative effects on the high expenditure in the administrative and academic resources. Predicting dropout has become an advantage for university administrators because it allows discovering students that are at risk of dropout as well as develop actions that allow taking decisions in a timely manner. This research presents a neural network approach through the application of multilayer perceptron algorithms and radial basis function. As input variables to the models, 11 factors were considered, which produce a negative influence in the desertion at the universities; the data was obtained from a survey of 2670 students of a Public University in Ecuador. The results showed that there is no significant difference in the accuracy rates of the proposed models which correspond to 96.3% for multilayer perceptron and 96.8% for radial basis function. As a conclusion, the studied models could be considered as an optimal option in terms of accuracy and concordance to predict dropout at the universities.

Index Terms—Prediction, university student desertion, neural networks, multilayer perceptron, radial basis function.

I. INTRODUCTION

Students' desertion at the universities is considered as a problem that affects higher education institutions worldwide [1]. Nowadays, the high dropout rates are considered as possible deficiencies in the undergraduate education system [2]. This can be evidenced in the academic and administrative management reports presented by government agencies worldwide. In the United States, the dropout rate at first and second year reaches 44.8% [3], 15.9% corresponds to the dropout rate in India [4], in countries such as Colombia, Ecuador and Brazil, the dropout rate at the universities exceeds 40% according to the United Nations (UN) in 2016.

As a result, the dropout is a consequence of a set of factors that interact with each other which have a negative influence in students to not finish their university studies successfully [5]. Dropout is considered as the voluntary or involuntary

early abandonment of a study program, in which the student discards the successful completion of their studies [6]. This phenomenon is still present in the higher education system and is related to negative effects such as the high economic and social cost that affects the student and the universities [7]. Although there is a large number of studies to try to solve the problem of dropout and determine its causes, there is a limited scientific production that incorporates machine learning algorithms such as neural networks that allows discovering the knowledge based on the nature of the variables obtained through the behavior of the students to know their condition of risk to dropout.

To design actions that allow to decrease the desertion rates at the universities, two models of neural networks, multilayer perceptron and radial base function are proposed. The results of this research will help universities administrators to promote changes in their academic policies and strategies in order to reduce their dropout rates.






This article is organized in five sections. Section II presents the background, theory and literature review. The materials and methods are presented in Section III. Section IV presents the results and the discussion and in Section V there is the conclusions.


II. BACKGROUND THEORY AND LITERATURE REVIEW


Several studies to predict dropout have been identified in the literature review, developed for the early identification of vulnerable and prone students to leave university classrooms. Table I shows some jobs that use neural networks classifier to predict dropouts in universities.

TABLE I. NEURAL NETWORKS TO PREDICT DROPOUT AT THE UNIVERSITIES


Description	References
Prediction of desertion in Turkey, based on data collected in an Information Technology Program.	[8]
Early warning system to predict the dropout of students in Information Literacy and Information Ethics careers.	[9]
Prediction of desertion in Spain	[10]
Prediction of the desertion through characteristics of the academic, financial, demographic performance of the students.	[11]
Prediction of the desertion in Czech Republic, through data collected in students of the career of Applied Informatics	[12]
Methodology to connecting the analysis of learning and	


medwell journals





 Redactar
 3 de 8 < > Es ⚙


Medwell Journals medwelljournals@gmail.com [a través de smtpsendemail.com](#)
dom., 9 dic. 2018 15:42 ☆ ↶ ⋮

para mí ▾

 inglés ▾ > español ▾ [Traducir mensaje](#)
[Desactivar para: inglés](#) x

Dear Ms. Mayra Alban

Thank you so much for providing us all required changes in your article, finally we are able to inform you that article is accepted for final publication and now it is pending only due to **non-payment**.

We request to you please generate your invoice from website and send us fee immediately to process it further.

With Best Regards

Dr. Muhammad Kamran
 Director Publications
Medwell Journals
 308-Lasani Town, Sargodha Road,
 Faisalabad, Pakistan
 Tel.: +92 41 500 3000
 Fax.: +92 41 881 5599
 e-mail: medwellonline@gmail.com
 URL: <http://www.medwelljournals.com>

Activar Windows

Ir a Configuración de PC para activar Windows