20th International Conference- Science, Technology and Innovation
Booklets

RENIECYT
Registro Nacional de Instituciones y Empresas Científicas y Tecnológicas
1702902 CONAHCYT

RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - Google Scholar DOI - REDIB - Mendeley - DIALNET - ROAD - ORCID - V|LEX

Title: Prediction model of the terminal efficiency of Computer Engineering at the Autonomous University of Tlaxcala applying Data Mining

Authors: SÁNCHEZ-SÁNCHEZ, Norma, MORA-LUMBRERAS, Marva Angélica, SÁNCHEZ-PÉREZ, Carolina Rocío and DÁVILA-GUTIÉRREZ, Blanca Leticia

Editorial label ECORFAN: 607-8695
BECORFAN Control Number: 2023-03
BECORFAN Classification (2023): 111213-0301

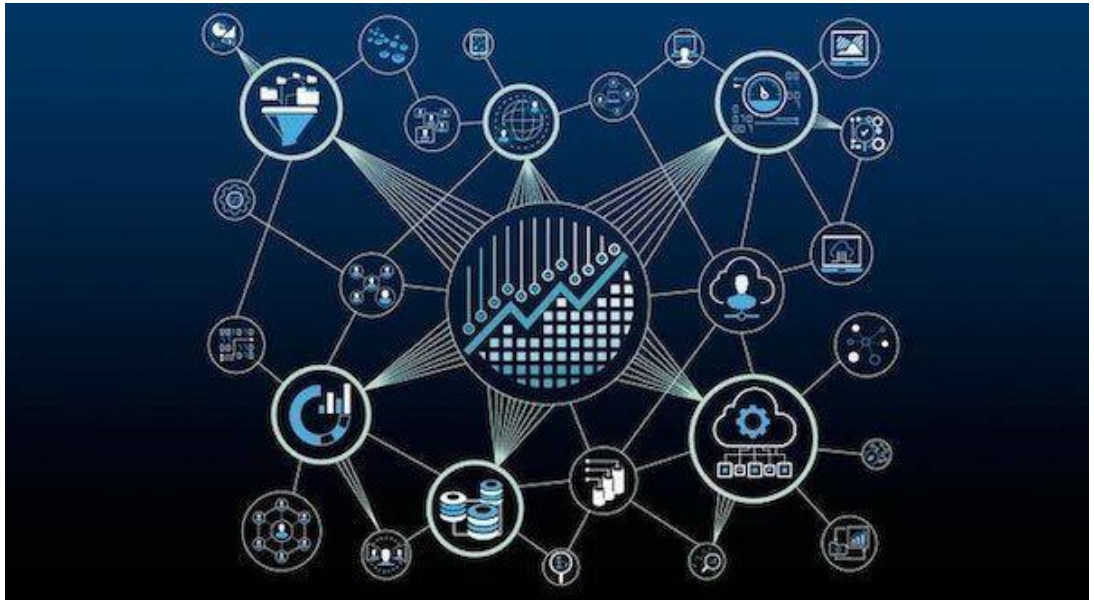
Pages: 17
RNA: 03-2010-032610115700-14

Contenido

- Introducción
- Metodología
- Resultados
- Conclusiones
- Referencias

Introducción

Hay un gran interés por aplicar las técnicas y métodos de Minería de Datos en los ambientes de la educación superior. Dentro de su aplicación esta la de transformar los datos en información útil para apoyar la toma de decisiones educativas. [1] [2]



Introducción

La minería de datos es una herramienta poderosa para predecir la eficiencia terminal en diferentes contextos educativos.

Permite analizar grandes volúmenes de datos recopilados de estudiantes y utilizar técnicas estadísticas y algoritmos de aprendizaje automático para identificar patrones, tendencias y factores clave que influyen en la eficiencia terminal.

Introducción

Trabajo	Indicadores a obtener	Fuente de Datos	Herramientas, técnicas y/o algoritmos de minería de datos	Metodología empleada
Un modelo basado en árboles de decisión para predecir la deserción estudiantil en la Educación Superior Privada. [3]	La deserción o el abandono en la Educación Superior Privada	Las base de datos de la Universidad César Vallejo – Lima Este	SPSS Clementine 12.0 , utilizando la técnica de minería de datos árboles de decisión	Metodología CRIPS-DM
Diseño de un Modelo predictivo aplicando Minería de Datos para identificar causas de Deserción Estudiantil Universitaria. [4]	Causas de Deserción Estudiantil	Sistema de Automatización de Información Integral de la IES	Algoritmos de selección de atributos aplicando el método de búsqueda (BestFirst) y atributo evaluador (CfsSubsetEval). Reglas de clasificación: JRIP, One R, ZeroR y Árbol de decisión J48 y REPtree. WEKA	Metodología de desarrollo del Modelo PredATISv1: ● Integración y recopilación. ● Selección, limpieza y transformación. ● Modelado Data Mining. ● Interpretación de resultados.

Introducción

Trabajo	Indicadores a obtener	Fuente de Datos	Herramientas, técnicas y/o algoritmos de minería de datos	Metodología empleada
Modelo basado en técnicas de minería de datos para análisis de factores de deserción estudiantil.[5]	Factores que influyen en la deserción	Base de Datos académicas de la Pontifica Universidad Javeriana	Programa SPSS Statistics	Metodología CRISP DM
Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos. [6]	Variables que influyen en la deserción	Data warehouse de la Universidad Arturo Prat, Chile	Modelos de clasificación: árboles de decisión, métodos bayesianos y redes neuronales. Herramienta WEKA	Metodología CRISP DM

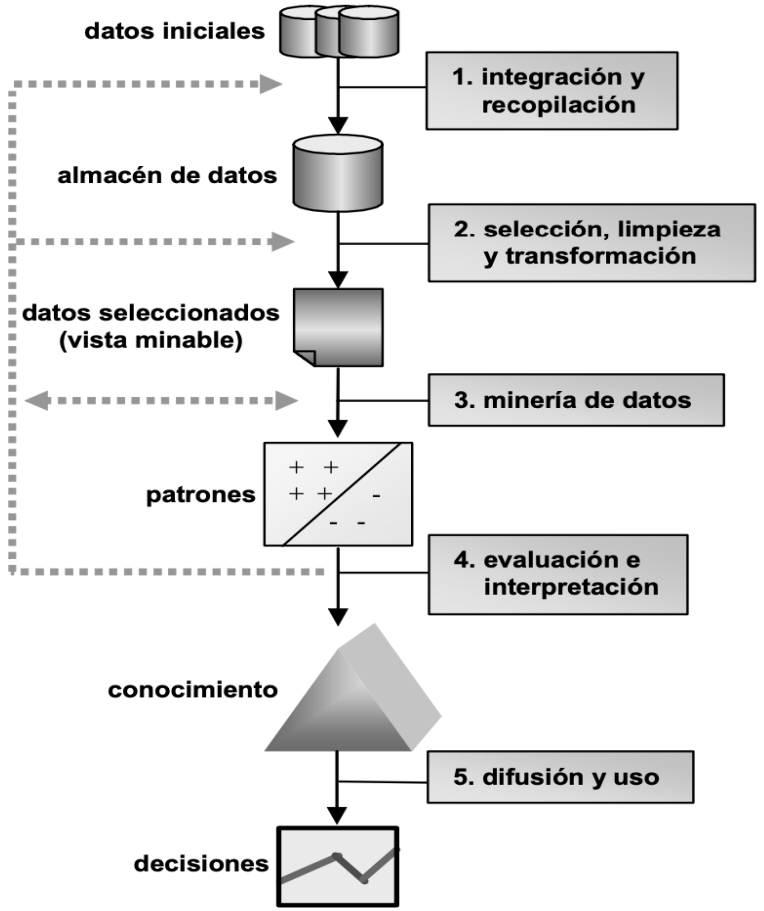
Minería de Datos

Las técnicas de minería de datos aplicadas a la tarea de la predicción, tienen como objetivo desarrollar un modelo que permita predecir el valor de la variable de entrada (variable dependiente) en función de un conjunto de variables predictoras (variables independientes).

El uso de la regresión lineal como un método simple es frecuentemente utilizado para la tarea de regresión, donde la regresión es también una tarea predictiva que consiste en aprender una función real que asigna a cada instancia un valor real. [7]

Metodología

Se hizo uso del proceso KDD por sus siglas en inglés (Knowledge Discovery from Databases), que define una metodología que provee una representación completa del ciclo de vida de un proyecto de Minería de Datos. [7]



Desarrollo
Integración y recopilación de datos

Recopilación de Datos.

Los datos se generan con una consulta SQL a la base de datos del Sistema Integral de información Administrativa de la Universidad Autónoma de Tlaxcala, y se exportan en un archivo CSV.

- *La muestra seleccionada corresponde a 2 cohortes generacionales, siendo de Otoño 2009 a Otoño 2010 de la Licenciatura en Ingeniería en Computación*

Desarrollo
Integración y recopilación de datos

Determinación de variables

De los datos recopilados, se examinaron 38 variables, seleccionando 21 variables representativas de los estudiantes, siendo las siguientes:

Genero, Estado Civil, Semestre, Semestre Concluido, Promedio por Semestre, Promedio Acumulado al Semestre, Materias Cursadas en el Semestre, Total de Materias Cursadas Acumuladas al Semestre, Materias Reprobadas en el Semestre, Total de Materias Reprobadas Acumuladas al Semestre, Materias Aprobadas en el Semestre, Total de Materias Aprobadas Acumuladas al Semestre, Creditos Cursados en el Semestre, Total de Creditos Cursados Acumulados al Semestre, Creditos Reprobados en el Semestre, Total de Creditos Reprobados Acumuladas al Semestre, Materias Aprobados en el Semestre, Total de Creditos Aprobados Acumulados al Semestre, Bajo Promedio, Alta Reprobacion y Rezago.

Desarrollo
Selección, limpieza y transformación

En esta fase, se genera la vista minable y patrones de entrada del entrenamiento:

- *Selección y Limpieza*
- *Transformación a variables*
- *Pre procesamiento de variables*

ID de la Variable	Descripción de la Variable
CLGENERO	Género 1 Masculino 2 Femenino
CLESTADOCIVIL	Estado civil 1 Soltero 2 Casado 3 Divorciado 4 Viudo 5 Union Libre
SEMESTRE	Semestre Cursado
Semestre_Concluido	0 No Concluido 1 Concluido
PROMEDIO_SEMESTRE	Promedio del semestre . Valores de 0 a 10
PROMEDIO_ACUMULADO_SEM	Promedio acumulado al semestre
MAT_CURSADAS_SEMESTRE	Número de materias cursadas del semestre
TOT_MAT_CURSADA_ACUM_SEM	Número total de materias cursadas acumuladas al semestre
MAT_REPROBADAS_SEMESTRE	Número de materias reprobadas del semestre
TOT_MATERIAS_REPROBADAS_SEM	Número total de materias reprobadas acumuladas al semestre
MAT_APROBADAS_SEMESTRE	Número de materias aprobadas del semestre
TOT_MATERIAS_APROBADAS_SEM	Número total de materias aprobadas acumuladas al semestre
CREDITOS_APROBADOS_SEMESTRE	Número de créditos aprobados del semestre
TOT_CRED_APROB_ACUM	Número total de créditos aprobados acumulados al semestre
CREDITOS_REPROBADOS_SEMESTRE	Número de créditos reprobados del semestre
TOT_CRED_REPROBADOS_ACUM	Número total de créditos reprobados acumulados al semestre
CREDITOS_CURSADOS_SEMESTRE	Número de créditos cursados del semestre
TOT_CRED_CURSADOS_ACUM	Número total de créditos cursados acumulados al semestre
BAJO_PROMEDIO	Bajo Promedio . Si el (PROMEDIO_ACUMULADO_SEM) > 7.5
ALTA_REPROBACION	Alta Reprobación Si ((MAT_REPROBADAS_SEMESTRE / MAT_CURSADAS_SEMESTRE) > 0.5) O ((TOT_MATERIAS_REPROBADAS_SEM/12) >0.6)
REZAGO	Rezago académico. Si (TOT_CRED_APROB_ACUM) < (TOT_CRED_CURSADOS_ACUM)

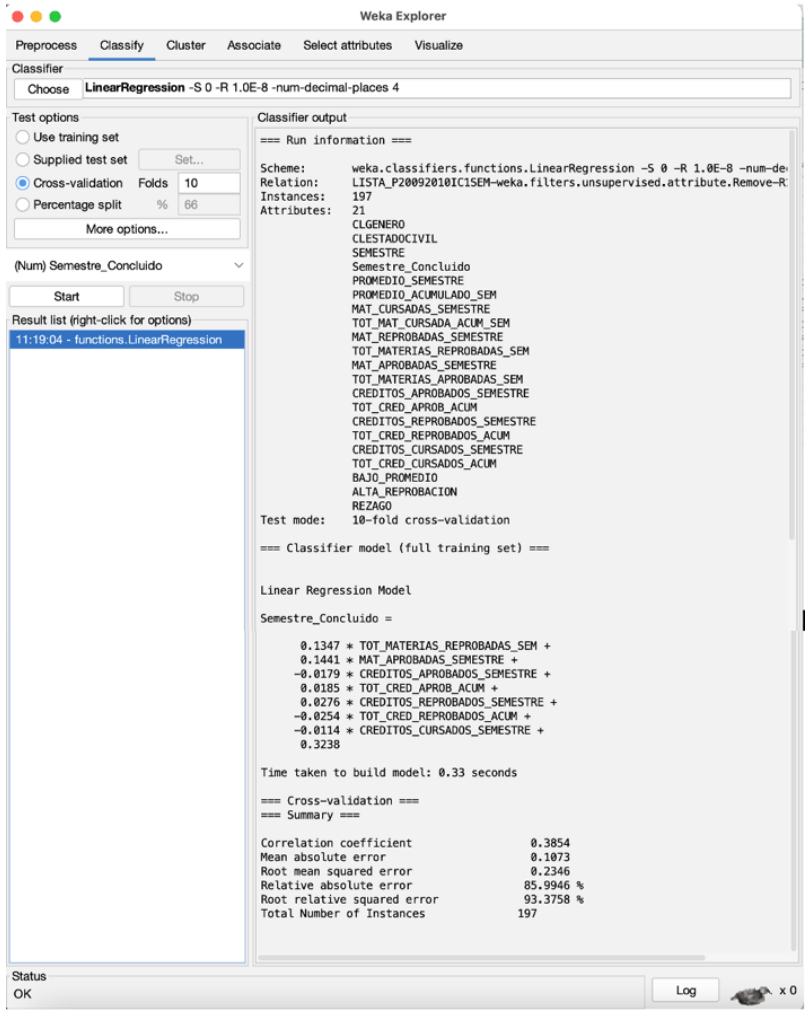
Desarrollo
Minería de datos

Para la aplicación de las técnicas de minería de datos, se uso el programa WEKA (Waikato Environment for Knowledge Analysis) desarrollado por la Universidad de Waikato de Nueva Zelanda. WEKA es un programa de uso libre (Licencia GNU) y está compuesto por un conjunto de algoritmos que implementan la mayoría de las técnicas de minería de datos. [10]

Desarrollo
Minería de datos

En Weka se aplica la técnica de minería de datos de regresión lineal sobre la variable Semestre_Concluido de los cohortes generacionales de Otoño 2009 a Otoño 2010 de la Licenciatura en Ingeniería en Computación para obtener el modelo predictivo para el primer semestre.

Al aplica la regresión lineal se obtiene el modelo con las variables predictoras que son estadísticamente significativas.



Desarrollo
Evaluación e interpretación

Con el modelo generado se realizaron las pruebas con la función obtenida a los cohortes generacionales de Otoño 2011, 2012, 2020, 2021 y 2022.

Se muestra el porcentaje de alumnos que concluyen el semestre. Se observa que de acuerdo al modelo se tiene una R de 85.99% y un R-cuadrado del 93.37% lo cual infiere un grado de confiabilidad estadística aceptable, lo cual se ve reflejado en los resultados obtenidos en la columna de Porcentaje de Acierto.

Cohorte	Total	Acertados por el modelo	Observada		Estimada	
			Sem Concluido	Sem No Concluido	Sem Concluido	Sem No Concluido
Otoño 2011	75	70	68	7	68	2
Otoño 2012	84	70	68	16	68	2
Otoño 2020	77	70	63	14	63	7
Otoño 2021	62	57	54	8	54	3
Otoño 2022	62	53	51	11	51	2
	344	320	304	56	304	16

Cohorte	Eficiencia de Sem Concluido (%)		
	Observada	Estimada	Porcentaje de Acierto
Otoño 2011	90.66	97.14	93.33
Otoño 2012	80.90	97.14	83.33
Otoño 2020	81.81	90.00	90.90
Otoño 2021	87.09	94.73	91.93
Otoño 2022	82.25	96.22	85.48

Resultados y Discusión

Se obtuvo un modelo que predice para el primer semestre del programa de Ingeniería en Computación si un alumno concluye o no el semestre. De acuerdo a las pruebas, la predicción del modelo es aceptable y puede usarse para estimar adecuadamente la conclusión del semestre de las cohortes que han cursado el primer semestre del programa educativo, siendo un punto de referencia para el semestre consecutivo.

Conclusiones

Las técnicas de minería de datos demuestran ser herramientas eficaces para obtener modelos que permiten predecir la eficiencia terminal.

Las variables identificadas en el modelo generado son variables relacionadas con los datos académicos del estudiante.

Al contar con el modelo se pueden tomar decisiones para diseñar los planes de desarrollo por parte de la alta dirección, al usar el modelo para estimar la eficiencia terminal, las estimaciones tendrán un sustento científico.

Referencias

[1] Heiner, C., Baker, R., & Yacef, K. (2006). Proceedings of Educational Data Mining workshop. 8th International Conference on Intelligent Tutoring Systems. pp. 250-257. https://www.educationaldatamining.org/EDM2008/uploads/proc/full%20proceedings.pdf

[2] Kabakchieva, D. (2013). Predicting Student Performance by Using Data Mining Methods for Classification. pp. 61- 72. https://doi.org/10.2478/cait-2013-0006

[3] Daza Vergaray, A. (2016). Un modelo basado en árboles de decisión para predecir la deserción estudiantil en la Educación Superior Privada. UCV-Scientia, 8(1), 59–73. https://doi.org/10.18050/RevUcv-Scientia.v8n1a7

[4] Maya Pérez, P. N., Aguilar C, J. R., Zamora R, R. A., & Barron A, J. M. (2018). Diseño de un Modelo predictivo aplicando Minería de Datos para identificar causas de Deserción Estudiantil Universitaria. STRATEGY, TECHNOLOGY & SOCIETY vol 7, 11-39. https://investigacion.upaep.mx/micrositios/cipu/assets/m1_16.pdf

[5] Bermúdez, S. C., Díaz, J. A. & Rodríguez, L. E. (2019). Modelo basado en técnicas de minería de datos para análisis de factores de deserción estudiantil. http://hdl.handle.net/10554/45510

Referencias

[6] Zarria, C., Arce, C., & Lam, J. (2016). Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos. Ciencia amazónica (Iquitos) 6, 73-84. https://doi.org/10.22386/ca.v6i1.110

[7] Orallo, J. H., Quintana, M. J. R., & Ramírez, C. F. (2004). Introducción a la minería de datos. Prentice Hall.

[8] Rodríguez, Y., y Díaz, A. (2009). Herramientas de Minería de Datos. Revista Cubana de Ciencias Informáticas, 3(3 - 4), 73 - 80. https://www.redalyc.org/articulo.oa?id=378343637009

[9] Herrero, J., y Molina, J. (2012). Técnicas de análisis de datos. Aplicaciones Prácticas utilizando Microsoft Excel y WEKA. Madrid, España: Universidad Carlos III. https://ocw.uc3m.es/pluginfile.php/4102/mod_page/content/10/data_mining_book.pdf

[10] Waikato University. (s.f.). Weka 3: Data Mining Software in Java. Isla Norte, New Zealand. https://www.cs.waikato.ac.nz/ml/weka/index.html.