

Resumen

La deserción estudiantil es un problema que se presenta en cada institución de educación superior, por lo cual en este trabajo de investigación se propone la aplicación de métodos y algoritmos de aprendizaje automático que permitieron la elaboración de un modelo capaz de realizar la estimación del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí. Con este objetivo se realizó el proceso de minería de datos a la información demográfica y académica de los estudiantes, contenida en el Sistema de Gestión Académica de la institución. Con la metodología CRISP DM se detallaron fases y procesos, tomando como muestra una carrera de cada una de las facultades durante el periodo académico (mayo 2014/febrero 2019). Luego se definió el criterio de inclusión para verificar la regularidad de los estudiantes y comprobar a través del coeficiente de correlación de Pearson la relación entre los estudiantes desertores y los no regulares, como estimador del riesgo de deserción. A través del proceso realizado se obtuvieron tres escenarios y se ejecutaron los algoritmos de regresión logística, KNN, redes neuronales, máquinas de vector de soporte y el algoritmo random forest que generó los resultados esperados y una mejor predicción sobre la deserción estudiantil. Por lo cual el mejor escenario fue durante los niveles 2,3 y 4 con un margen de error de 0.05 valorado por las métricas de evaluación y existiendo una alta correlación en cada una de ellas, en el área bajo la curva de 0.95 y un Puntaje F1 de 0.95.

PALABRAS CLAVES:

- **DESERCIÓN ESTUDIANTIL**
- **APRENDIZAJE AUTOMÁTICO**
- **MINERÍA DE DATOS**
- **CRISP DM**

Abstract

Student desertion is a problem that occurs in each higher education institution. This piece of research proposes the application of methods and algorithms of automatic learning that allowed the development of a model capable of estimating the risk of desertion in students at Universidad Técnica de Manabí. For this purpose, the process of data mining was carried out on the demographic and academic information of the students, found at the Academic Management System in this higher institution with the CRISP DM methodology. Phases and processes were detailed, taking as a sample a career of each of the faculties during the academic period (May 2014 / February 2019). Then, the inclusion criterion was defined to verify the regularity of the students and to check the relationship between the dropout and non-regular students, as an estimator of the risk of desertion, through the Pearson correlation coefficient. Through the process carried out, three scenarios were obtained and the logistic regression, KNN, neural networks, support vector machines and the random forest algorithm were executed generating the expected results and a better prediction of students' desertion. Therefore, the best scenario was at levels 2, 3 and 4 with a margin of error of 0.05, measured by the evaluation metrics and existing a high correlation in each of them in the area under the curve of 0.95 and an F1 Score of 0.95.

KEYWORDS:

- **STUDENT DESERTION**
- **MACHINE LEARNING**
- **DATA MINING**
- **CRISP DM**