

## Resumen

En el ámbito educativo es primordial la toma de decisiones, en décadas pasadas se realizaban cuando finalizaba el año escolar, actualmente con la tecnología actual se puede predecir o tener una herramienta que nos ayude a mejorar la calidad de la educación en cualquier época del año lectivo. La presente tesis realiza recoge la información académica y social de los alumnos del Colegio Particular Aristóteles Bilingüe con esta información se realiza un proceso de limpieza de datos, para luego evaluar 3 modelos de predicción que son: regresión logística, Random Forest y Árbol de decisión. Construidos estos modelos se los evalúa mediante la matriz de confusión, la curva de ROC (Característica Operativa del Receptor) y el coeficiente de Kappa. Siendo el árbol de decisión el escogido por tener 92% de exactitud y 0,74 de coeficiente de Kappa que lo vuelve el más sustancial de acuerdo con el grado de concordancia de los valores precedentemente nombrados. Elegido el modelo de árbol de decisión se procede a ejecutar y guardar la predicción de que los alumnos 4to año aprueben o reprobren el año escolar. Además, se construye un Tablero de Inteligencia de Negocios con información relevante a la institución y también un apartado con información relevante a la proyección que se realizó en el modelo propuesto.

*Palabras clave:* minería de datos educativos, modelos de predicción, predicción de rendimiento académico, deserción escolar, tablero institucional escolar de inteligencia de negocios.

## Abstract

In the educational field it is essential to make decisions, in past decades they were made at the end of the school year, nowadays with current technology we can predict or have a tool to help us improve the quality of education at any time of the school year. This thesis collects the academic and social information of the students of the Colegio Particular Aristóteles Bilingüe with this information a process of data cleaning is carried out, to then evaluate 3 prediction models which are: logistic regression, Random Forest and Decision Tree. Once these models are built, they are evaluated by means of the confusion matrix, the ROC (Receiver Operating Characteristic) curve and the Kappa coefficient. Being the decision tree the one chosen for having 92% accuracy and 0.74 Kappa coefficient which makes it the most substantial according to the degree of agreement of the values previously mentioned. Once the decision tree model has been chosen, the prediction that 4th grade students will pass or fail the school year is executed and stored. In addition, a Business Intelligence Board is built with information relevant to the institution and a section with information relevant to the projection made in the proposed model.

*Keywords:* educational Data Mining, prediction models, academic performance prediction, school dropout, business intelligence school institutional board.