

## **RESUMEN**

El índice de deserción estudiantil en el Ecuador es un problema que afecta a la sociedad, a la economía y a la Universidad por lo tanto este proyecto se enfoca en el desarrollo de un modelo que permita predecir la posibilidad que tiene un alumno de reprobado una materia o finalmente desertar, de tal forma que se pueda crear estrategias y tomar decisiones que lo solventen utilizando la minería de datos que es un conjunto de técnicas que permiten descubrir patrones de comportamiento a partir de un gran conjunto de datos. Se recuperó la información personal y académica de los alumnos de la Universidad de las Fuerzas Armadas – ESPE Matriz durante los últimos cinco años. Para el desarrollo del proceso de minería de datos se ha utilizado la metodología CRISP-DM que está constituida por seis etapas: Comprensión del Negocio, Comprensión de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue. En las dos primeras etapas se utilizó las herramientas Pentaho Data Integration para la creación del Data Warehouse y R Studio para la exploración estadística de los datos, en la tercera etapa se utilizó Open Refine y Pentaho Data Integration y finalmente en las dos últimas etapas la herramienta WEKA que permite la extracción del conocimiento utilizando distintas técnicas de minería de datos.

Palabras clave:

- **DESERCIÓN UNIVERSITARIA**
- **MINERÍA DE DATOS.**
- **CRISP-DM**
- **WEKA.**
- **PENTAHO DATA INTEGRATION.**
- **R STUDIO.**

## **ABSTRACT**

The dropout rate in Ecuador is a problem that affects society, the economy and the university. Therefore, this project focuses on the development of a model that allows predicting the possibility of a student failing a subject. Finally Deserting, in such a way that you can create strategies and make decisions that solvent using data mining is a set of techniques that allow you to detect patterns of behavior from a large dataset. The personal and academic information of the students of the Universidad de las Fuerzas Armadas - ESPE Matriz during the last five years was recovered. For the development of the data mining process, the CRISP-DM methodology has been used, which consists of six stages: Business Understanding, Data Comprehension, Data Preparation, Modeling, Evaluation and Deployment. In the first two stages we used the Pentaho data integration tools for the creation of the data warehouse and R Studio for the statistical exploration of the data, in the third stage we used Open data integration of Pentaho and Pentaho in the Two last stages of the WEKA tool that allow the extraction of knowledge using various techniques of data mining.

Keywords:

- **UNIVERSITY DROPOUTS**
- **DATA MINING.**
- **CRISP-DM**
- **WEKA.**
- **PENTAHO DATA INTEGRATION.**
- **R STUDIO.**