

RESUMEN

Este estudio propone el análisis y minería de datos de los bachilleres y profesionales del Ecuador; y su relación con las actividades económicas, aplicado al Sector del Conocimiento y Talento Humano del Ecuador desde el año 1993 al 2015”. Dicho estudio se realizó mediante la recopilación de datos del repositorio central de la entidad rectora, misma que se creó con el objetivo de desarrollar políticas públicas, que mejoren la gestión de conocimiento y la matriz productiva del país. Para el desarrollo de la investigación se utilizó la metodología CRISP-DM, por sus ventajas y flexibilidad a la hora de elaborar el diseño y construcción. Para la implementación del estudio se utilizó las herramientas de software libre R-Studio, Pentaho Data Integration y Postgresql. En la minería de datos se usó el algoritmo de clustering k-means y el algoritmo de asociación a priori. Entre los principales hallazgos y resultados está la generación dinámica de cinco clústers que muestran la relación entre el sector la educación y las actividades económicas, a los que se les aplicó las reglas de asociación que generó nuevo conocimiento útil para la toma de decisiones, que está descrito en el capítulo cuatro.

PALABRAS CLAVE

- **MINERÍA DE DATOS**
- **CRISP-DM**
- **R-STUDIO**
- **ASOCIACIÓN**
- **CLUSTERING**

ABSTRACT

This study proposes the analysis and mining of data of high school graduates and professionals of Ecuador; and its relationship with economic activities, applied to the Knowledge and Human Talent Sector of Ecuador from 1993 to 2015, this study was carried out through the collection of data from the central repository of the governing entity, which was created with the objective of to develop public policies that improve knowledge management and the productive matrix of the country. For the development of the research, the CRISP-DM methodology was used, due to its advantages and flexibility when it comes to design and construction. The free software tools R-Studio, Pentaho Data Integration and Postgresql were used to implement the study. In data mining, the k-means clustering algorithm and the a priori association algorithm were used. Among the main findings and results is the dynamic generation of five clusters that show the relationship between the education sector and economic activities, to which the rules of association that generated new knowledge useful for decision-making, which is described in chapter four.

KEY WORDS

- **DATA MINING**
- **CRISP-DM**
- **R-STUDIO**
- **ASSOCIATION**
- **CLUSTERING**