



ANÁLISIS DE LOS MÉTODOS DE AGRUPAMIENTO Y CLASIFICACIÓN EN SQL SERVER

**ELGUEDA LEZANA CHRISTIAN
SEPÚLVEDA HIRZER ELIZABETH
INGENIERO COMERCIAL**

RESUMEN

En esta investigación se examinan los registros de la base de datos que contienen los campos que describen el tipo de seguros adquirido por un cliente de una entidad financiera, determinado agrupamientos y clasificaciones, a través del algoritmo EM (Expectation Maximization), incorporado en SQL Server a través de su herramienta Análisis Services, para analizar el comportamiento de esta herramienta.

El análisis realizado se debe a que faltan criterios concretos para interpretar con claridad los resultado de Agrupamientos y Clasificación en SQL Server, ya que el software no entrega información detallada de los resultados que se obtienen al aplicar una técnica sobre un conjunto de datos.

De los resultados obtenidos, la ruta del nodo no es un criterio estándar para el método agrupamiento, sin embargo, SQL server entrega las probabilidades asociadas a los escenarios de los grupos formados, la cual sirve para conocer qué porcentaje del total de registros se encuentran en determinado grupo, y que por lo tanto se puede utilizar como criterio para el agrupamiento de datos. Para la clasificación, la ruta del nodo si es un criterio estándar valido para la clasificación de nuevos datos, ya que no se presentan las anormalidades encontradas en el agrupamiento de datos. Además mediante la discretización de los datos, la clase también entrega otro criterio para la clasificación, debido a que se sabe en que rango fluctúan los registros de las hojas del Árbol de decisión para la clasificación.