



UNA APLICACIÓN CONJUNTA EN LA EXPLORACIÓN DE DATOS ACADÉMICOS.

Paula Zuliani , A. Cechich.
Departamento de Informática y Estadística
Universidad Nacional del Comahue
{pzuliani, acechich}@uncoma.edu.ar
Buenos Aires 1400, 8300 Neuquen, ARGENTINA.

Introducción.

La minería de datos colecciona, almacena y organiza datos para ser usados en diferentes aplicaciones tales como medicina, educación, manufacturación y procesos de control. El éxito de la minería de datos requiere la integración de algunas tecnologías donde los sistemas de administración de bases de datos juegan un rol fundamental. Estos sistemas son usados para organizar y proveer acceso efectivo a grandes cantidades de datos.

La planificación de un proyecto de minería de datos comienza con la integración y depuración o transformación de las fuentes de datos, la exploración de los datos, la revisión y refinamiento de los resultados de la exploración, y el reporte de los resultados finales[1]. Por lo general se implementa alguna acción para determinar si los resultados de la minería de datos son poderosos o no.

Material y métodos.

Un proceso de clasificación usa información histórica para predecir el funcionamiento o los miembros de una clase dada[2]. La base de datos con información histórica, también denominada *base de datos de entrenamiento*, contiene registros con algunos atributos. Un atributo designado es llamado el *atributo dependiente*, y los otros son llamados *atributos predictores*. El objetivo es construir un modelo que tome los atributos predictores como entradas y asignarle un valor al atributo dependiente.

C4.5 es el nombre colectivo para un conjunto de programas de computación que construyen modelos de clasificación para el descubrimiento y análisis de patrones encontrados en los registros de la base de datos.

Resultados.

- ✓ La adquisición y el procesamiento de los datos juegan un rol importante.
- ✓ La calidad –afectada por la cantidad de valores desconocidos y cambiados– hace que el clasificador no sea confiable ni lo suficientemente exitoso en la exploración de los datos.
- ✓ La falta de consideración de atributos con importante poder predictivo para la clasificación hace que los resultados de la clasificación no sean satisfactorios–el modelo no es confiable, no trabaja bien, etc.

Conclusiones.

La exploración de datos erróneos provoca únicamente una pérdida de esfuerzo y tiempo.



Referencias Bibliográfica.

- [1] Bhavani Thuraisingham, *A Primer for Understanding and Applying Data Mining*, IEEE IT Professional, Vol. 2, N° 1, (Feb. 2000), 28-31
- [2] Venkatesh Ganti, Johannes Gehrke, and Raghuram Ramakrishnan, *Mining Very Large Databases*, Communications of the ACM, Vol. 32, N° 8, (August 1999), 38-45.