



### Jesús Luis García Garrido,

CEO DE AVANSIS Y MIEMBRO DEL GRUPO DE DESARROLLO DE LA SOCIEDAD DE LA INFORMACIÓN DE AUTELSI

# La evolución de las técnicas de **Data Mining**

**E**n la actualidad el Data Mining sigue basándose en técnicas EDA (Exploratory Data Analysis) para explorar los datos, aplicando algoritmos estadísticos con cierta implementación de inteligencia artificial en base a reglas de autoaprendizaje limitada y rígida. La complejidad de la adecuada aplicación de estas metodologías en base a la adaptación forzada a las herramientas existentes dificulta la obtención de unos resultados realmente útiles y flexibles.

La próxima generación de herramientas evoluciona las tecnologías aplicadas, aprovechando la potencia de las redes neuronales artificiales, manejando una evolucionada interpretación eficiente a través de árboles de decisión dotados de reglas de inducción flexibles y aplicando algoritmos genéticos.

Las redes neuronales artificiales utilizan modelos predictivos no lineales para adquirir el conocimiento a través del autoaprendizaje. Los algoritmos aplicados están diseñados para “pensar”, actuar y tomar decisiones con un razonamiento más cercano al humano. Estos modelos resultan altamente complejos de utilizar, incluso por los expertos, ya que generan los resultados a través de modelos predictivos difíciles de gestionar. Aplicando reglas de inducción conseguimos automatizar el conocimiento y el aprendizaje de forma no supervisada. Normalmente las reglas de inducción se utilizan con más éxito en bases de datos con alta cardinalidad, que permiten generar los patrones adecuados para realizar una mejor predicción.

La aplicación de árboles de decisión como técnica de Data Mining se basa en automatizar reglas para clasificar conjuntos de datos. Partiendo de un nodo de inicio, y en función de las decisiones, este se puede dividir en dos o más nodos y se replica la algorítmica. Es una técnica rápida ya que requiere menos pre-tratamiento de los datos y muy práctica para la exploración y para la predicción,

por la aplicación de reglas excluyentes y exhaustivas.

Los algoritmos genéticos se basan conceptualmente en la genética y la selección natural, en la combinación y la mutación. Se utilizan en el reconocimiento de patrones, ya sea buscando la clasificación o como una herramienta de optimización.

Los conceptos, metodologías, técnicas y tecnologías de Data Warehouse y Data Mining están ya presentes de forma habitual en múltiples áreas de negocio para analizar las grandes bases de datos de forma eficiente. Hay infinitas utilidades para cualquier sector, y tienen un importante futuro asociadas a las aplicaciones de negocio, ya que aportan nuevas posibilidades de análisis, predicción automatizada de las tendencias y comportamientos en una gran base de datos en tiempo prácticamente real y sin una férrea dependencia de equipos técnicos altamente cualificados. Son ya imprescindibles en áreas como ventas y marketing, la gestión sanitaria, las utilidades financieras y muchos más, aportando un gran beneficio en estas a través del análisis rápido de los datos, imprescindible para una excelente calidad del proceso de toma de decisiones. ♦

**Las redes neuronales artificiales utilizan modelos predictivos no lineales para adquirir el conocimiento a través del autoaprendizaje**