

Cos'è il Data Mining

È il processo di estrazione di conoscenza da banche dati di grandi dimensioni tramite l'applicazione di algoritmi che individuano le associazioni "nascoste" tra le informazioni e le rendono visibili.

In altre parole, col nome data mining si intende l'applicazione di una o più tecniche che consentono l'esplorazione di grandi quantità di dati, con l'obiettivo di individuare le informazioni più significative e di renderle disponibili e direttamente utilizzabili nell'ambito del decision making.

L'estrazione di conoscenza (informazioni significative) avviene tramite individuazione delle associazioni, o "patterns", o sequenze ripetute, o regolarità, nascoste nei dati. In questo contesto un "pattern" indica una struttura, un modello, o, in generale, una rappresentazione sintetica dei dati.

Il termine data mining è utilizzato come sinonimo di knowledge discovery in databases (KDD), anche se sarebbe più preciso parlare di knowledge discovery quando ci si riferisce al processo di estrazione della conoscenza, e di data mining come di una particolare fase del suddetto processo (la fase di applicazione di uno specifico algoritmo per l'individuazione dei "patterns").

Per una descrizione delle fasi del KDD si veda:

- **Il processo di estrazione di conoscenza**

U.Fayyad, G.Piatetsky-Shapiro, P.Smyth, R.Uthurusamy ("Advances in knowledge discovery and data mining") definiscono il KDD come "nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data". Questa definizione consente di mettere in luce l'aspetto inferenziale del processo, nonché le caratteristiche dei "patterns" in termini di validità, novità (non già noti), potenziale utilità e comprensibilità.

Altre definizioni

"Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques." (Gartner Group).

"Data mining is the exploration and analysis, by automatic and semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules." (M.J.A. Berry, G. Linoff).