

**PROGRAMMA DEL CORSO DI  
TECNICHE INFORMATICHE PER L'ANALISI ECONOMICA**

**A.A. 2003/2004**

*Prof. Donato Malerba*

**Premessa.** Negli ultimi anni è cresciuto il volume dei dati economici, finanziari e di marketing memorizzati nelle basi di dati. Alcuni esempi sono i survey condotti da enti governativi, le transazioni di borsa, le informazioni sulle vendite nelle grandi catene di supermercati. Il denominatore comune di tutti questi esempi è che la mole dei dati da elaborare richiede l'uso di strumenti informatici specializzati per l'analisi dei dati. Il Data Mining è il processo di analisi di dati memorizzati nelle basi di dati e mirante a scoprire e estrarre conoscenza in modo più automatizzato delle classiche tecniche. Il processo è "guidato dai dati" nel senso che si pone minore enfasi su ipotesi a priori di quanto si faccia con le classiche tecniche statistiche. Approcciando un'analisi come un problema di ricerca di elementi di conoscenza, piuttosto che come il test di un'ipotesi, si mira a scoprire nuove relazioni o "pattern" nei dati.

**Obiettivi.** Il corso presenta i concetti e le metodologie evolute di business intelligence e di data mining, con riferimento a problemi decisionali di natura economica.

**Prerequisiti:** statistica, basi di dati.

**Programma del corso a.a. 2003-2004**

1. Architetture e modelli per l'analisi dei dati.

Sistemi informativi e sistemi informatici. Il ruolo del sistema informativo nell'organizzazione: la rappresentazione per funzioni, la rappresentazione per processi, la rappresentazione di Anthony. Dati operazionali e dati decisionali. Tecnologie di Business Intelligence. Sistemi di supporto alle decisioni (DSS), Executive Information Systems (EIS) e Management Information Systems (MIS). Caratteristiche di un data warehouse. Architettura di un data warehouse. Schema del data warehouse: schema a stella, a costellazione e schema a fiocco di neve. OLAP e operazioni per l'analisi dei dati: drill down e roll up. ROLAP e MOLAP.

2. Scoperta di conoscenza nelle basi di dati.

La scoperta di conoscenza nelle basi di dati: definizione e problemi. Il processo della scoperta di conoscenza nelle basi di dati: la selezione, il preprocessing, la trasformazione, il data mining, l'interpretazione dei risultati, la valutazione dei risultati. I task di data mining. Algoritmi di data mining per compiti di classificazione: alberi di decisione, K-nearest-neighbour (K-NN). Algoritmi di data mining per compiti di regressione: funzioni di regressione, alberi di regressione, alberi di modelli. Algoritmi di data mining per compiti descrittivi: regole di associazione. Valutazione dei risultati con il metodo holdout, cross-validation, bootstrap. Metodo di bagging per migliorare la precisione del classificatore.

3. Esercitazioni

Il sistema KDB2000: architettura software, input, output e algoritmi.

## ***Principali testi di riferimento***

Paolo Giudici

*Data Mining – metodi statistici per le applicazioni aziendali*

McGraw-Hill, 2001

Capitoli: 1, 2.1, 2.2, 4.3, 4.9, 5.1, 6.5, 12

Marco Tagliavini, Aurelio Ravarini, Donatella Sciuto

*Sistemi per la gestione dell'informazione*

Apogeo, 2003

Capitoli: 1

Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees, Alessandro Zanasi

*Discovering Data mining: from concepts to implementation*

Morgan Kaufmann, 2001

Capitoli: 3, 4

Paolo Atzeni, Stefano Ceri, Piero Fraternali, Stefano Paraboschi, Riccardo Torlone

*Basi di dati: architetture e linee di evoluzione*

McGraw-Hill, 2003

Capitoli: 8

Le trasparenze presentate durante il corso sono sul sito web

<http://www.di.uniba.it/~malerba/tixae/>