DATA MINING

Ce este Data Mining?

- Extragerea cunostintelor din seturi mari de date
- Extragerea aurului este numita “*gold mining”,* si nu *“rock or sand mining*”
- Astfel, data mining ar fi corect numita “knowledge mining from data”, dar s-a optat pentru varianta mai scurta

Exemplu 1

- Aproximativ 13 milioane de clienti contacteaza in fiecare luna call center-ul celei mai mari banci din America, Bank of America
- In trecut, fiecare persoana asculta aceeasi reclama, indiferent daca era relevanta sau nu pentru interesele sale
- Din dorinta de a fi cat mai relevanti pentru fiecare client si avand acces la profilul fiecarui client, au facut astfel incat o persoana care suna sa asculte o reclama de interes
- **Tehnicile de data mining ajuta la identificarea celei mai potrivite abordari de marketing pentru fiecare client, bazandu-se pe profilul individual al fiecarui cleint**

Exemplu 2

- Fostul presedinte Bill Clinton, a mentionat ca la putin timp dupa evenimentele din 11 septembrie 2001, agentii FBI au examinat foarte multe baze de date si i-au gasit pe 5 dintre teroristi in aceste baze de date
- Unul dintre teroristi avea 30 carti de credit cu o datorie de aprox $250,000 si venise in tara de mai putin de 2 ani
- Un alt terorist, Mohammed Atta, avea declarate 12 adrese diferite dintre care doar 2 erau reale
- Clinton a concluzionat ca ar trebui sa se faca o cautare proactiva in acest tip de date pentru a identifica din timp potentiali teroristi

Ce este Data Mining?

- “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.” *(The Gartner Group, www.gartner.com)*

- “Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner”. *(David Hand, Heikki Mannila, and Padhraic Smyth, Principles of Data Mining, MIT Press, Cambridge, MA, 2001.)*

- “Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases”. (*Peter Cabena, Pablo Hadjinian, Rolf Stadler, JaapVerhees, and Alessandro Zanasi, Discovering Data Mining: From Concept to Implementation, Prentice Hall, Upper Saddle River, NJ, 1998.)*

Cat de raspandit este data mining-ul?

STATISTICS INTERN: SEPTEMBER–DECEMBER 2003

Work with Basketball Operations

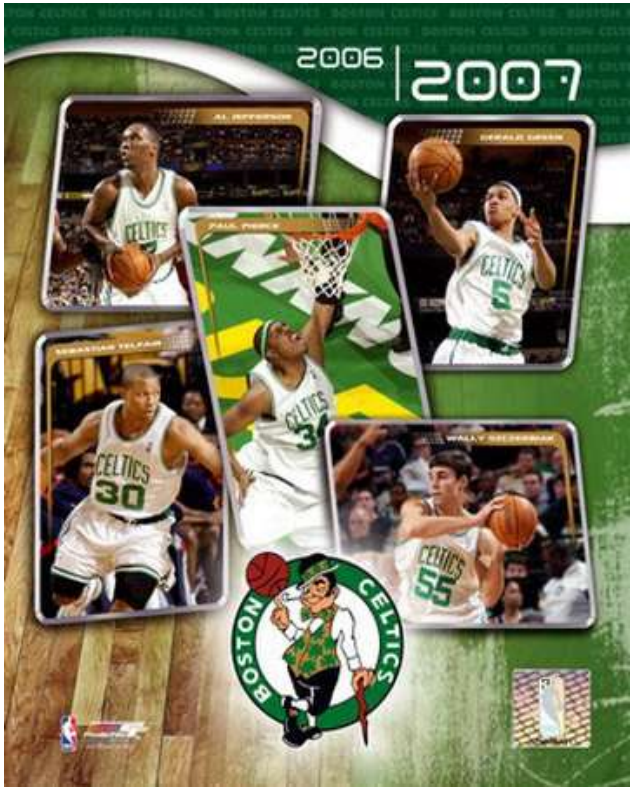
Responsibilities include:

- Compiling and converting data into format for use in statistical models
- Developing statistical forecasting models using regression, logistic regression, **data mining**, etc.
- Using statistical packages such as Minitab, SPSS, XLMiner

Experience in developing statistical models a differentiator, but not required.

Candidates who have completed advanced statistics coursework with a strong knowledge of basketball and the love of the game should forward your résumé and cover letter to:

Boston Celtics
Director of Human Resources
151 Merrimac Street
Boston, MA 02114



The Boston Celtics, http://www.nba.com/celtics/

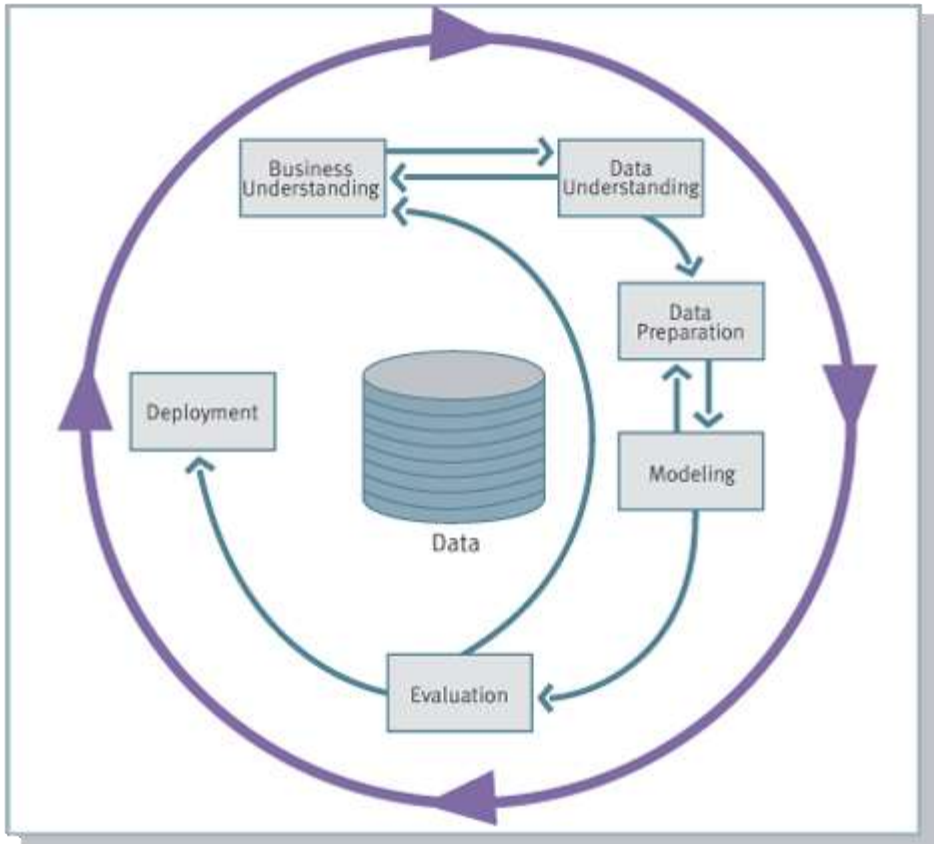
De ce Data Mining?

- cresterea masiva a colectiilor de date
- folosirea depozitelor de date
- accesul la o multitudine de date pe web
- cresterea puterii calculatoarelor si a posibilitatilor de memorare a datelor

CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING

CRISP-DM

an industry- and tool-neutral data mining process model.



www.crisp-dm.org

CRISP–DM: The Six Phases

- *Business understanding phase*
- *Data understanding phase*
- *Data preparation phase*
- *Modeling phase*
- *Evaluation phase*
- *Deployment phase*

www.crisp-dm.org

Ce se poate face cu Data Mining?

- Descriere
- Estimare
- Predictie
- Clasificare
- Clustering
- Asociere