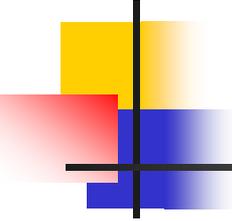


Data Mining

Shahram Hassas

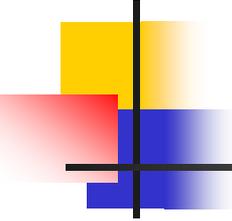
Math 382

Professor: Shapiro



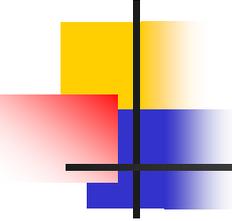
Agenda

- Introduction
- Major Elements
- Steps/ Processes
- Examples
- Tools used for data mining
- Advantages and Disadvantages



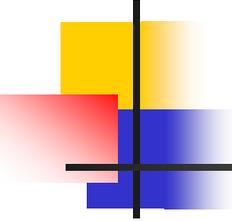
What is Data Mining?

- Described as the method of comparing large volumes of data looking for more information from a data
- Data Mining is the process of analyzing data from different perspectives and summarizing it into useful information which can be used to increase revenue, and cut costs.



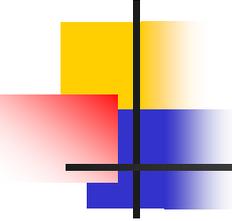
What is Data Mining?

- Data mining is the results of Classical statistics, artificial intelligence and machine learning.
- Classical statistic plays the main role in data mining. Artificial intelligence applies human interest processing to statistical problems.



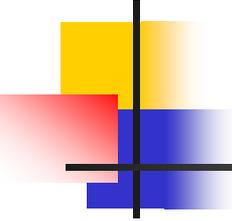
What is it used for?

- To assist in the analysis of collections of observations and finding correlations among all of the fields and relationship between the facts in databases.



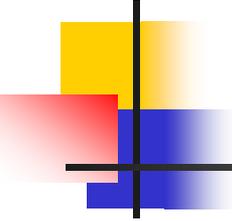
Examples

- grocery chain stores discovered that when men bought diapers on Thursdays and Saturdays, they also attempt to buy beer. It is also showed that these shoppers typically did their weekly grocery shopping on Saturdays and only bought a few items on Thursdays. The grocery chain used this newly discovered information in various ways to increase revenue for instance; they moved the beers closer to the diapers and, they could make sure beer and diapers were sold at full price on Thursdays.



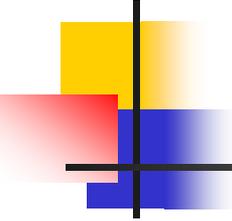
Data mining consists of five major elements:

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.



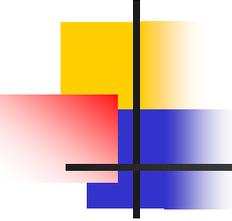
Data Mining Goal

- Data mining application goals are predictions, and it allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships by exploration to estimate the expected result(s).



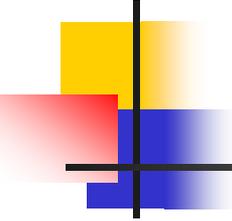
Examples

- The National Basketball Association (NBA) is exploring a data mining application that can combine and analyze the records of basketball games by using advanced Scout software. This program helps to find patterns derived from game statistics, images, and the movements of the players themselves.



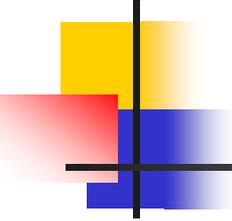
3 Steps Data Mining Process

- **Stage 1: Exploration.** This stage usually starts with data preparation which may involve cleaning data, data transformations, selecting subsets of records
- **Stage 2: Model building and validation.** This stage involves considering various models and choosing the best one based on their predictive performance
- **Stage 3: Deployment.** That final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimates of the expected outcome



Some of the tools used for data mining are:

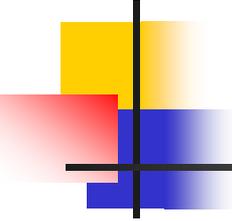
- Artificial neural networks - Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Decision trees - Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
- Rule induction - The extraction of useful if-then rules from data based on statistical significance.
- Genetic algorithms - Optimization techniques based on the concepts of genetic combination, mutation, and natural selection.
- Nearest neighbor - A classification technique that classifies each record based on the records most similar to it in an historical database.



Reasons for the growing popularity of Data Mining

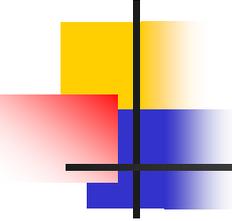
- **Growing Data Volume**
- **Limitations of Human Analysis**
- **Cost and efficiency compared to other Statistical data applications**

ADVANTAGES OF DATA MINING



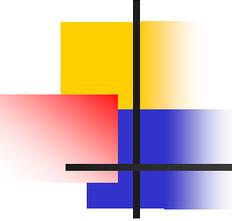
- Data mining has been used in area of science and engineering, such as bioinformatics, genetics, medicine, education, agricultural, electrical power engineering, and law enforcement

ADVANTAGES OF DATA MINING



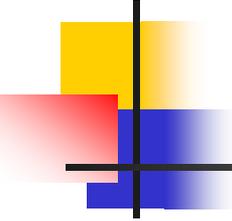
- In the field of human genetics, usage of data mining attempts to find out how the changes in an individual's DNA sequence affect the risk of developing common diseases such as cancer.

ADVANTAGES OF DATA MINING



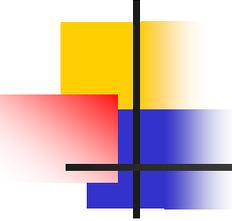
- **Law enforcement:** Data mining can aid law enforcers in identifying criminal suspects.
- **Electrical power Engineering:** data mining techniques have been used for condition monitoring of high voltage electrical equipment to obtain valuable information on the insulation's health status of the equipment

DISADVANTAGES OF DATA MINING

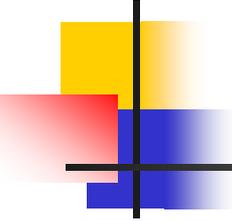


- **Privacy Issues:** For example, according to the Washington Post, in 1998, CVS had sold their patient's prescription purchases to a different company
- American Express also sold their customers' credit card purchases to another company.

DISADVANTAGES OF DATA MINING Cont...



- **Security issues:** Although companies have a lot of personal information about us available online, they do not have sufficient security systems in place to protect that information.
- **Misuse of information:** Some of the companies will answer your phone call based on your purchase history. If you have spent a lot of money or bought a lot of products from one company, your call will be answered really soon. So you should not think that your call is really being answered in the order in which it was received.



References:

http://en.wikipedia.org/wiki/Data_mining

[http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm/data_mining.htm,](http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm/data_mining.htm)

[http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace.](http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace) Palace, Bill. "Data Mining."

<http://www.niehs.nih.gov/research/resource/databases/gac/guid.cfm>