# DATA MINING RESEARCH: RETROSPECT AND PROSPECT

Prof(Dr).V.SARAVANAN
&
Mr. ABDUL KHADAR JILANI
Department of Computer Science
College of Computer and Information Sciences
Majmaah University
Kingdom of Saudi Arabia

# AGENDA

1. **Introduction**
2. **The Top 10**
   - **Data Mining Applications**
   - **Challenging Research Problems in Data Mining**
   - **Data Mining Algorithms**
   - **Data Mining Keywords**
   - **Mistakes in Data Mining**
   - **Data Mining Software**
   - **Data Mining Researchers**
   - **Data Mining Authors**
   - **Universities/Research Institutions**
   - **Data Mining Companies**
   - **Data Mining Conferences**
   - **Data Mining Journals**
3. **Controversies**

# Data Mining

- New buzzword, old idea.
- Inferring new information from already collected data.
- Traditionally job of Data Analysts
- Computers have changed this.
  Far more efficient to comb through data using a machine than eyeballing statistical data.

# Data Mining vs. Data Analysis

- In terms of software and the marketing thereof Data Mining != Data Analysis
- Data Mining implies software uses some intelligence over simple grouping and partitioning of data to infer new information.
- Data Analysis is more in line with standard statistical software (ie: web stats).

# Sources of Data for Mining

- Databases (most obvious)
- Text Documents
- Computer Simulations
- Social Networks

# The Top 10 Data Mining Applications

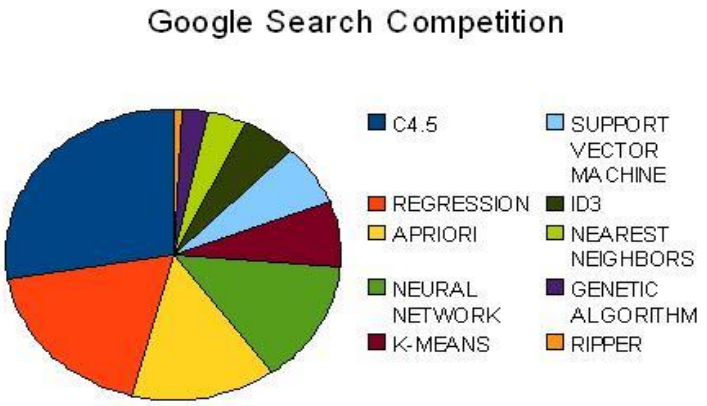
1. Social Networking
2. Health Care
3. Banking
4. Insurance
5. Telecommunication
6. Education
7. Marketing
8. Sports
9. Advertisement
10. Bio Medical

# The Top 10 Challenging Research Problems in Data Mining

1. Simultaneous mining over multiple data types
2. Over-fitting vs. not missing the rare nuggets
3. Sequential and Time Series Data
4. Mining Complex Knowledge from Complex Data
5. Data Mining in Graph Structured Data (Social Networking)
6. Distributed Data Mining and Mining Multi-agent Data
7. Data Mining for Biological and Environmental Problems
8. Automated Data Mining
9. Security, Privacy and Data Integrity
10. When to use which algorithm?

# The Top 10 Data Mining Algorithms

1. C4.5 ALGORITHM
2. REGRESSION ALGORITHM
3. APRIORI ALGORITHM
4. NEURAL NETWORK ALGORITHM
5. K-MEANS ALGORITHM
6. SUPPORT VECTOR MACHINE ALGORITHM
7. ID3 ALGORITHM
8. NEAREST NEIGHBORS ALGORITHM
9. GENETIC ALGORITHM
10. RIPPER ALGORITHM




*Source: http://mydatamine.com/*

# The Top 10 Data Mining Key Words

1. Data Mining
2. Social Network
3. Large Scale
4. Machine Learning
5. Information Retrieval
6. Indexation
7. Search Engine
8. Cluster Algorithm
9. Web Search
10. Web Pages

*Source: www.kdnuggets.com*

# The Top 10 Mistakes in Data Mining

1. Focus on training
2. Rely on one technique
3. Ask the wrong question
4. Listen (only) to the data
5. Accept leaks from the future
6. Discount pesky cases
7. Extrapolate
8. Answer every inquiry
9. Sample casually
10. Believe the best model

*Source: http://datamininglab.com*

# The Top 10 Data Mining Software - Licensed

1. IBM SPSS Modeler
2. SAS Data Mining
3. Angoss Knowledge Studio
4. Microsoft Analysis Services
5. Oracle Data Mining
6. Think Analytics
7. Viscovery
8. Portrait
9. IBM DB2 Intelligent Miner
10. Statistica Data Miner

*Source: http://www.predictiveanalyticstoday.com/*

# The Top 10 Data Mining Software - Free

1. KNIME
2. R
3. ML-Flex
4. Databionic ESOM tools
5. Orange
6. Natural Language Tool Kit (NLKT)
7. SenticNet API
8. ELKI
9. Rapid Miner
10. SCaViS

*Source: http://www.predictiveanalyticstoday.com/*

# The Top 10 Data Mining Researchers

Dean Abbott

Michael Berry

Tom Davenport

John Elder

Rayid Ghani

Anthony Goldbloom

Vincent Granville

Gregory Piatetsky-Shapiro

Karl Rexer

Eric Siegel

***Source: http://www.deep-data-mining.com/***

# The Top 10 Data Mining Authors

1. Jiawei Han, H-Index=69
2. Philip S. Yu, 47
3. Rakesh Agrawal, 46
4. Christos Faloutsos, 39
5. Heikki Mannila, 36
6. Eamonn J. Keogh, 35
7. George Karypis, 35
8. Jian Pei, 34
9. Padhraic Smyth, 34
10. Hans-Peter Kriegel, 33

*Source:http://www.quora.com/*

# The Top 10 Universities

1. Carnegie Mellon University
2. Massachusetts Institute of Technology
3. Stanford University
4. University of California—Berkeley
5. University of Illinois—Urbana-Champaign
6. Cornell University
7. University of Washington
8. Princeton University
9. Georgia Institute of Technology
10. University of Texas—Austin

*Source: http://grad-schools.usnews.rankingsandreviews.com/*

# Top 10 Companies

1. Actian
2. Birst
3. BllomReach
4. CBIG Consulting
5. Cirro
6. Digital Reasoning
7. Flutura Solutions
8. Fractal Analytics
9. Hadapt
10. Link Analytics

***Source: www.kdnuggets.com***

# The Top 10 Data Mining Conferences

1. KDD - Knowledge Discovery and Data Mining
2. ICDE - International Conference on Data Engineering
3. CIKM - International Conference on Information and Knowledge Management
4. ICDM - IEEE International Conference on Data Mining
5. SDM - SIAM International Conference on Data Mining
6. PKDD - Principles of Data Mining and Knowledge Discovery
7. PAKDD - Pacific-Asia Conference on Knowledge Discovery and Data Mining
8. WSDM - Web Search and Data Mining
9. DASFAA - Database Systems for Advanced Applications
10. ICWSM - International Conference on Weblogs and Social Media

*Source: www.kdnuggets.com*

# The Top 10 Data Mining Journals

1. TKDE - IEEE Transactions on Knowledge and Data Engineering
2. IPL - Information Processing Letters
3. VLDB - The Vldb Journal
4. DATAMINE - Data Mining and Knowledge Discovery
5. Sigkdd Explorations
6. CS&DA - Computational Statistics & Data Analysis
7. Journal of Knowledge Management
8. WWW - World Wide Web
9. Journal of Classification
10. INFFUS - Information Fusion

*Source: http://academic.research.microsoft.com*

# Data Mining Controversies

- Your data is already being mined, whether you like it or not.
- Many web services require that you allow access to your information [for data mining] in order to use the service.
- Google mines email data in Gmail accounts to present account owners with ads.
- Facebook requires users to allow access to info from non-Facebook pages.

- **Facebook's Beacon Advertising program**

**What Beacon does:**
**"when you engage in consumer activity at a [Facebook] partner website, such as Amazon, eBay, or the New York Times, not only will Facebook record that activity, but your Facebook connections will also be informed of your purchases or actions."**

***Source: http://trickytrickywhiteboy.blogspot.com/2007/11/beware-of-facebooks-beacon.html***

# Top 10 Recommended Resources and Works Consulted

1. www.kdnuggets.com
2. http://academic.research.microsoft.com
3. http://grad-schools.usnews.rankingsandreviews.com
4. http://www.quora.com/
5. http://www.predictiveanalyticstoday.com/
6. http://datamininglab.com
7. http://mydatamine.com/
8. http://www.deep-data-mining.com/
9. http://www-01.ibm.com/software/analytics/spss/products/modeler/
10. http://kdl.cs.umass.edu/papers/jensen-neville-nas2002.pdf

# THANK YOU!