

Data Mining To Boost Your Business

Faisal Kamran



INFORMATION TECHNOLOGY
UNIVERSITY

Outline

- **Motivation for data mining**
- **Data mining techniques**
- **Applications in business**
- **Ethics of data mining**
- **Conclusion**



Motivation for Data Mining: the Data Flood

- **Huge amounts of data are available in digital form**
 - Internet
 - IP Traffic logs
 - Scientific data
 - Customer profiles
 - ...
- **No longer possible to analyze the data “manually”**



We are flooded by data but starving for information

Jiawei Han



INFORMATION TECHNOLOGY
UNIVERSITY

Some Examples of Data Size

- **155 million websites on the Internet**
- **WhatsApp handles more than 50 billion messages daily (phonearena.com)**
- **294 billion emails sent per day (Radicati group, 2010)**



What is Data Mining?

- Data mining is the use of automatic techniques to “discover *knowledge*”
 - Data driven discovery
 - Making implicit knowledge explicit
 - Collecting data, Preprocessing, Mining, Visualizing, ...



1110000001000000000000110000100000110000000000000010000000
000111000001110111011100000110011100000101100001000000000
110000000100000001000011001000000011000000000000010000100
11100000010000001000001100000000001010000000000010000000
111000000100000000000011000000000011000000000000010000000
010111000000110111011100100110011100011111100001000000000
000111000000110101011100000110011101011111100001000000000
111000000100000000000011000010000011001000000000010100000
110000000100000000000011000000000011000000000000010000000
111100101100001000000011111000100011000000011000110000000
000111000000110111011100000110011100011111000001000000000
000111011011110111111100000111111000111111100111001100000
0000001000000011000000111110000000000000001011000010000000
000000111011110011000000000001100001100000000111001100000
000000111011110011100000000001100000100000000111001100000
000100001011110011100001000001100000100010000111001100000
000000011010110010110110000001100010100000000111001100000
000000100000001000010101111000000000000000011000110000000
000000100000001000000011111000000000000000011000110000000
000000100010001000000011111000000000000000010000111000000
001000100000001000000010111000000000000000011000110000000
000111000000110111011100000110011100011111100001000000000
000111000000110111011100000110011100011111100001000010000
000000100000001000000011111000000000000000011000110000000

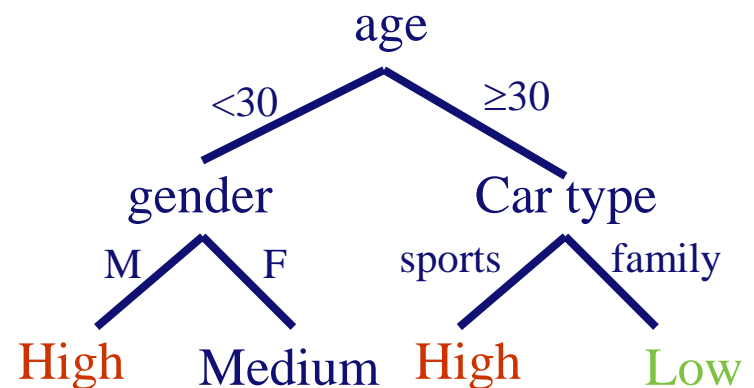
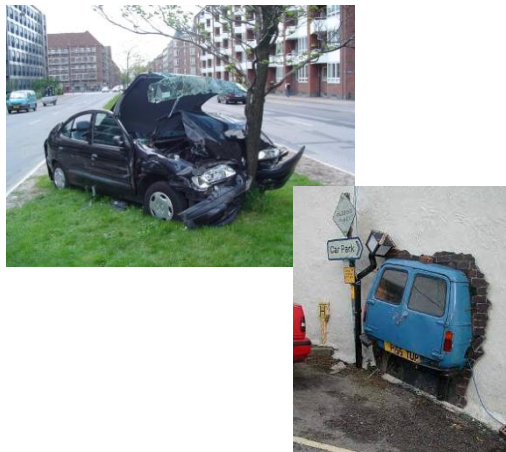
11110111100000000000000000000000100000000001000000000000000000000000
111111111000
1111111110000000000000000000100010000000000000000000000000000000100000
111011111000
1111111111111111111000000010000000000000000000000000000010010000000
00000011111111111100
00000011111011111000100000000000000100000000000000000000000000000000
00000010111111110000000000010100000000000000000000000000000000000000
000000111111111100
00000011111110111000
000100110111111100
000000000000000000111111111111111111111111000000000000000000000000
000000000000000000111111111111111111111111000000000000000000000000
000000000000000000110111110011111111111111100000000100000000000000
001000000100000001111111111111111111111110000000000000000000000000
000001000000000000111111111111111111110111100000000000000000000000
000000000000000000111111101111111111111110000000000000000000000000
000000000000000000111101111111111111111111000000000000000000000000
000000000000000000111111111111111101111111111111111111100000000000
0000010000001000
0000000000000000001000
0000000000000000001000
000000001000
0000100100
0000000000000000001000



Data Mining Techniques: Classification

- Learn a *model* based on *labeled* data.
- The model can be used for *prediction*.

Example:



Data Mining Techniques: Clustering

- **Example:**

 [Advanced Search](#)
[Preferences](#)

[George W. Bush - Wikipedia, the free encyclopedia](#)

Open-source encyclopedia article provides personal, business and political information about the President, his policies, and public perceptions and ...

en.wikipedia.org/wiki/George_W_Bush - 459k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Bush \(band\) - Wikipedia, the free encyclopedia](#)

Bush was a post-grunge band from the UK, formed in 1992. Their debut album was the self-released Sixteen Stone in 1994. They have sold well over 10 million ...

[en.wikipedia.org/wiki/Bush_\(band\)](http://en.wikipedia.org/wiki/Bush_(band)) - 60k - [Cached](#) - [Similar pages](#) - [Note this](#)

[More results from en.wikipedia.org »](#)

[President of the United States - George W. Bush](#)

The Oval Office contains speeches and statements of President **Bush**, a description of policy priorities, biographies, and photo essays.

www.whitehouse.gov/president/ - 21k - [Cached](#) - [Similar pages](#) - [Note this](#)

[More results from www.whitehouse.gov »](#)

[Gavin Rossdale: gavinrossdalefans.com](#)

The former lead singer of **BUSH**, the platinum selling alt rock juggernaut, Gavin can now be seen UP CLOSE at this intimate Past Show. ...

gavinrossdalefans.com/ - 38k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Bush Furniture, Inc](#)

Bush designs and manufactures quality, ready to assemble, entertainment centers, TV stands, home office and business furniture.

www.bushfurniture.com/ - 26k - [Cached](#) - [Similar pages](#) - [Note this](#)



Data Mining Techniques: Pattern Mining

- Find regularities, trends, patterns that frequently occur in the data



Data Mining to Boost Your Business

- Mining customer data
 - Which customers are likely to “churn”?
 - Concentrate on these customers
 - What kind of customers like a specific offer?
 - Which promotions to offer to a customer?
 - Capital One has been able to cut the company’s losses from fraud by more than 50% by deploying data mining methods.
 - Products people with a similar profile like
 - How to retain loyal customers?

Data Mining to Boost Your Business

- **Increased sales**

Safeway (UK)'s data mining system found that the top - spending 25% customers often purchase a particular cheese product ranked below 200 in sales. Normally, without the data - mining results, the product would have been discontinued and would disappoint the best customers. Safeway continues to order this cheese, although it is ranked low in sales.



Data Mining to Boost Your Business

- **Less expenditures**
- **Bank of America identified savings of \$4.8 million in 2 years by using a credit risk management system, i.e., examination of only borderline applicants.**
- **BBC's data mining based program scheduler determines the timing to show programs as good as the best planner but at much less cost.**



Data Mining to Boost Your Business

- **Effective decision making**

Bell Atlantic developed telephone technician dispatch system. They must decide what type of technician to dispatch to resolve the reported complain. Bell Atlantic save more than 10 million dollars per year by using data mining rule based system because they make fewer erroneous decisions.



Data Mining in Policy Framing

- **Metro Bus Transit System (MBTS)**
 - to predict the frequency of buses at certain times
 - optimal routes for future expansion
 - fare estimation
- **Online college admission system**
 - Identification of poor performing institutions
 - Estimation of future intakes well in advance



Data Mining in Policy Framing

- **Law and order:**
 - **Learn characteristics of people that should be under surveillance**
 - **Find patterns in crimes; e.g., sudden increase in burglaries in one particular area**



Ethics of Data Mining: Careful Analysis

- **Berkeley case (1973)**

	Applicants	Admitted
Men	8442	44%
Women	4321	35%

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%



Ethics of Data Mining: Privacy Issues

- **First step in data mining:**
 - **Collect data**
 - **Buying history (Loyalty cards)**
 - **Browsing behavior**
 - **Buy the data from a company**
 - **Survey data; e.g. CBS (average income/postal code)**
- **Combining different sources may lead to privacy issues**
 - **Information asymmetry**



Conclusion

- **Data mining = using automatic techniques to find patterns in data**
- **Many useful applications:**
 - **Automatic model building for future decision making**
 - **Increased revenue for companies**
 - **More efficient policing**
- **However, also dangers!**
 - **Misleading findings**
 - **Privacy issues**



Acknowledgement

- **Special thanks to Dr. Toon Calders (ULB), Dr. Saeed Ul Hassan (ITU) for the support.**
- **Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition. Mehmed Kantardzic. 2011 by Institute of Electrical and Electronics Engineers. Published 2011 by John Wiley & Sons, Inc.**



Data Mining and Business Intelligence Group at ITU



Dr. Faisal Kamran (ITU)

**PhD, Eindhoven University of
Technology, The Netherlands**

Email: faisal.kamiran@itu.edu.pk



Dr. Saeed Ul Hassan (ITU)

**PhD, Asian Institute of Technology,
Thailand**



Dr. Asim Karim (LUMS)

PhD, Ohio State University, USA



INFORMATION TECHNOLOGY
UNIVERSITY

Thank You for Your Attention!



INFORMATION TECHNOLOGY
UNIVERSITY