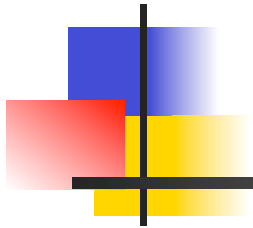


Data Mining

INFS 4203/7203



Dr. Mohamed A. Sharaf
School of Information Technology & Electrical
Engineering

Semester 2, 2018



Today...

- You!
- Me!
- Dupree!
 - INFS4203/7203



You

-
- Who are you?
 - What should you do?

Quick Facts about the 2015 Offering



Course	Enrollment
INFS4203 (Undergraduates)	13
INFS7203 (Postgraduates)	45

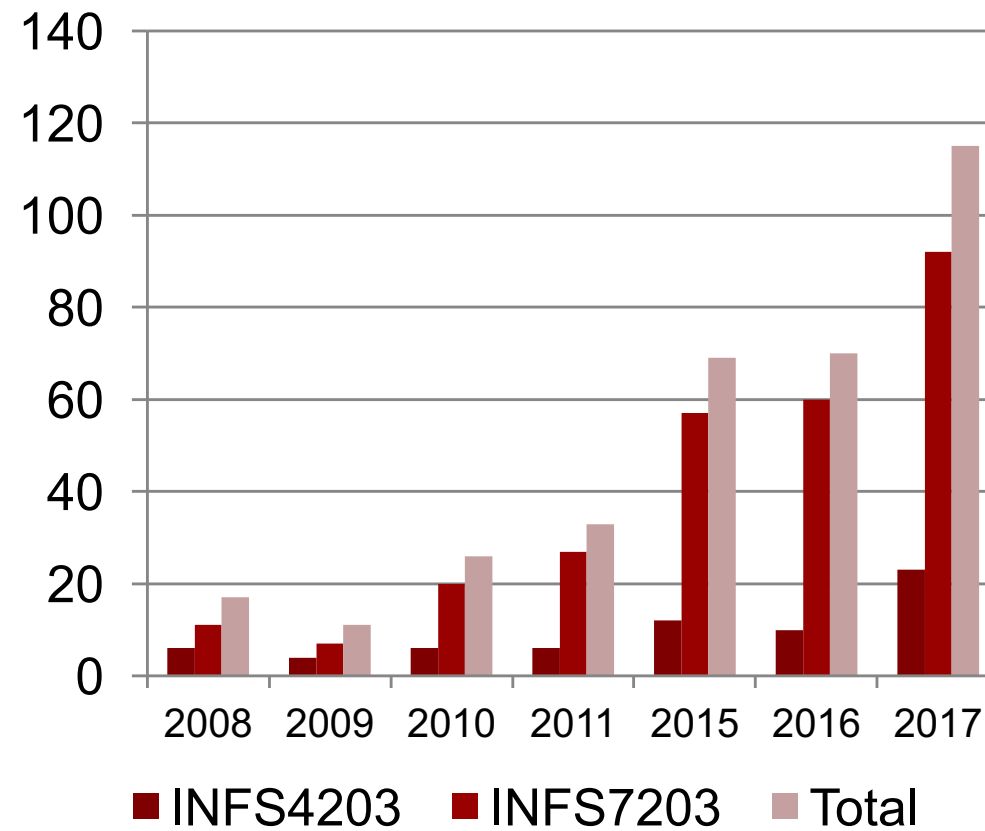
■ Multiple Programs:

- Master/Grad Cert./Grad Diploma in:
 - computer science, information technology, engineering science, bioinformatics, geog info systems, economics
- Bachelors in:
 - engineering, information technology, science
- Exchange students

Quick Facts about the 2018 Offering

Course	Enrollment
INFS4203 (Undergraduates)	33
INFS7203 (Postgraduates)	174

- **Multiple Programs:**
 - Master/Grad Cert./Grad Diploma in:
 - computer science, information technology, data science, engineering science, bioinformatics, GIS, economics
 - Bachelors in:
 - engineering, information technology, science, computer science
 - Exchange students/ study abroad



You

-
- Who are you?
 - What should you do?

Learning Activities



	Lectures	Tutorials
When	Wednesday 8:00-9:50am	4 sessions
Where	01-E109	67-142/342

- Tutorials:
 - Start in week 3

Material



- Lecture Notes
- Recommended Readings
 - Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Introduction to Data Mining. 1st Edition. 2006.
 - J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2001

Course Web Site

- All announcements and materials for **both** INFS4203/7203 will be on the course blackboard site
- Please check at least once a week

Assessments



Assessment	Percentage
Midterm Exam – 19 Sept.	30%
Project – 9 Oct.	20%
Final Exam	50%

Today...

- You!
- Me!
- Dupree!
 - INFS4203/7203



Teaching Staff

- **Course Coordinator:**

- Dr. Mohamed A. Sharaf

Email: m.sharaf@itee.uq.edu.au

Room: 78-649

- **Tutors:**

- Mr Oluseun Aremu
- Ms Dan He

UQ INFS Courses

- INFS1200/7900 Information Systems
- **INFS2200/7903 Relational Database Systems**
- INFS3200/7905 Advanced Database Topics
- INFS3202/7202 Web Information Systems
- INFS3208/7208 Cloud Computing
- **INFS4203/7203 Data Mining**
- INFS4205/7205 Adv. Tech. for High Dimensional Data
- INFS7410 Information Retrieval and Web Search

INFS courses are offered by the:

Data and Knowledge Engineering Research Group in ITEE
one of the strongest database research groups in the world
(<http://www.itee.uq.edu.au/~dke>)

INFS2200/7903

- In-depth understanding of Database Management Systems (DBMS) technology
 - System-Oriented
 - Understand the functions of a DBMS
 - Learn how these functions are implemented
 - Gain practical experience in applying this knowledge using commercial DBMS

INFS1200 vs INFS2200



INFS1200



INFS2200

INFS4203/7203

- In this course:
 - Data mining concepts and techniques for discovering **interesting** data **patterns** hidden in **large amount of data**
 - Algorithms for supporting **scalable and efficient** data mining



INFS4203/7203

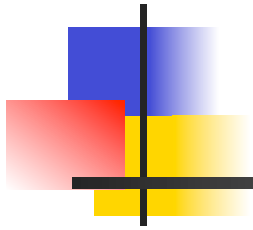


Today...

- You!
- Me!
- Dupree!
- INFS4203/7203



Introduction to Data Mining

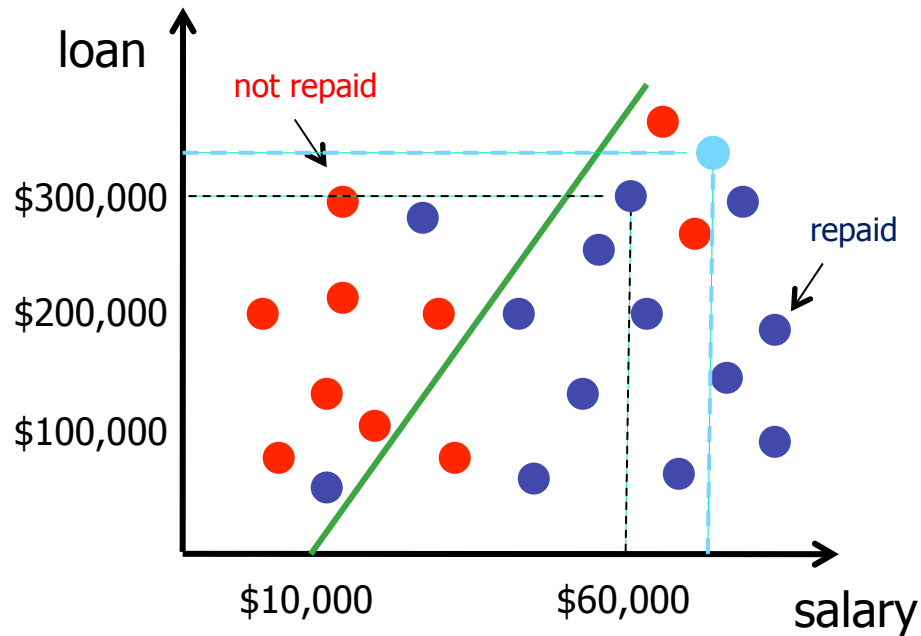


INFS4203/7203

- In this course:
 - Data mining concepts and techniques for discovering **interesting** data **patterns** hidden in **large amount of data**
 - Algorithms for supporting **scalable and efficient** data mining



Example: Credit Risk

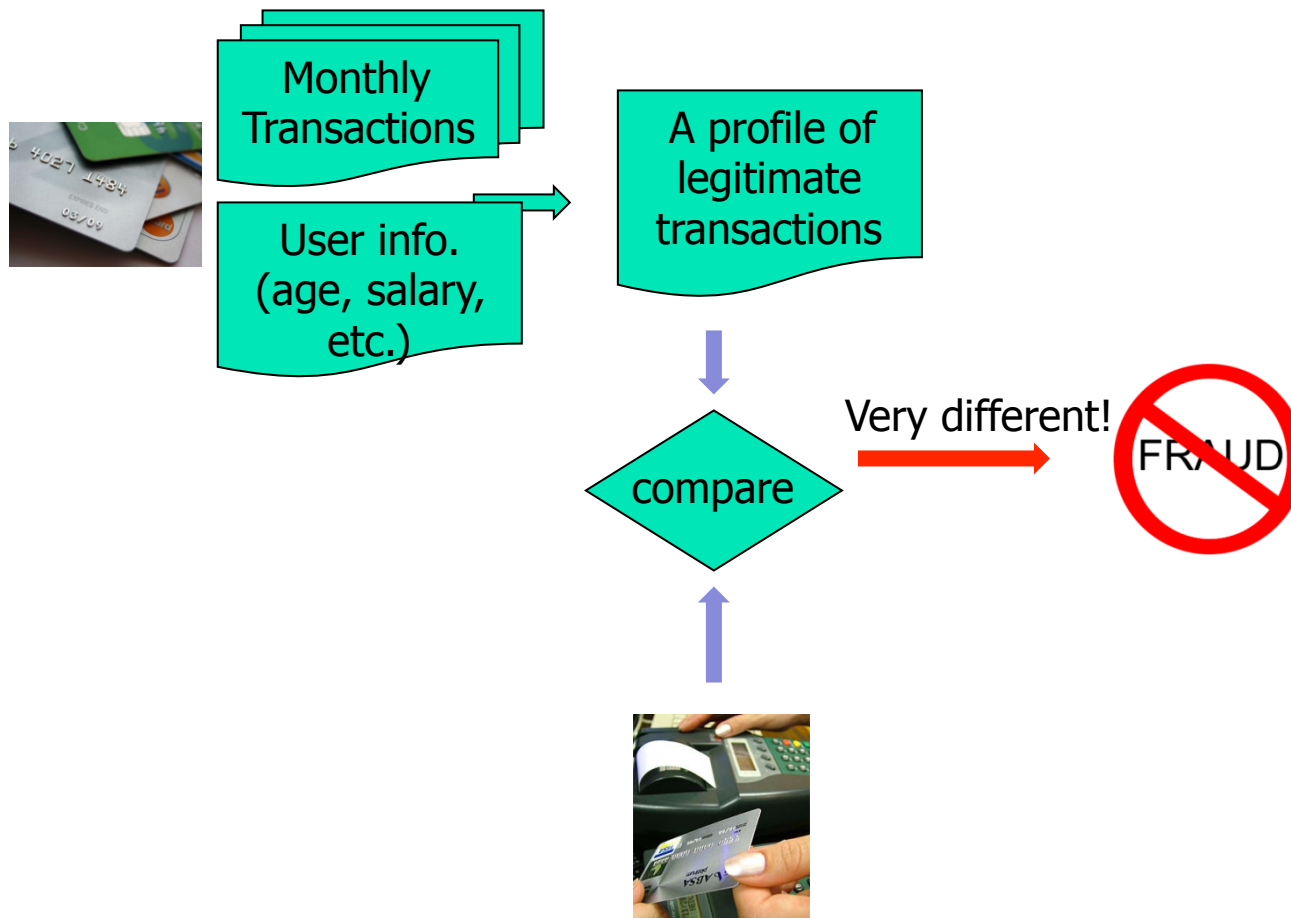


Given: a loan application

Task: decide if the bank should approve that loan

Idea: use data from previous loans!

Example: Credit Card Fraud Detection



Example: Market Basket Analysis

Identify **potential cross-selling opportunities** among related items



Anything interesting?

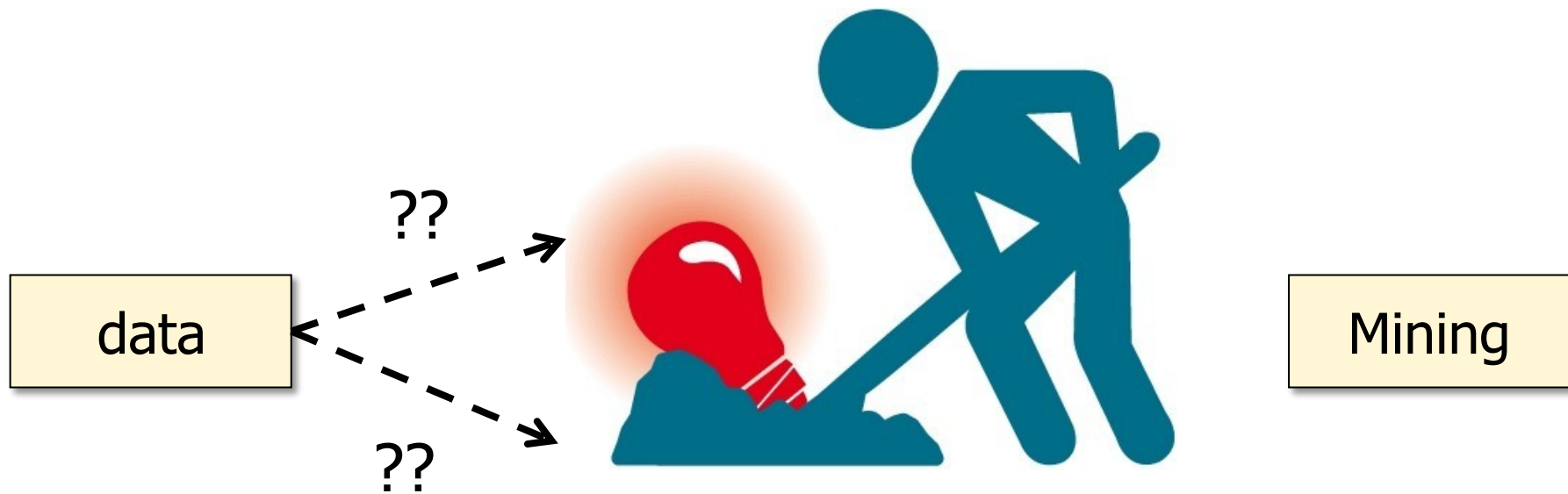
Bread → Milk (100%)
Diapers → Milk (100%)
Diapers → Beer (66%)

Customers who buy diapers also tend to buy beer!

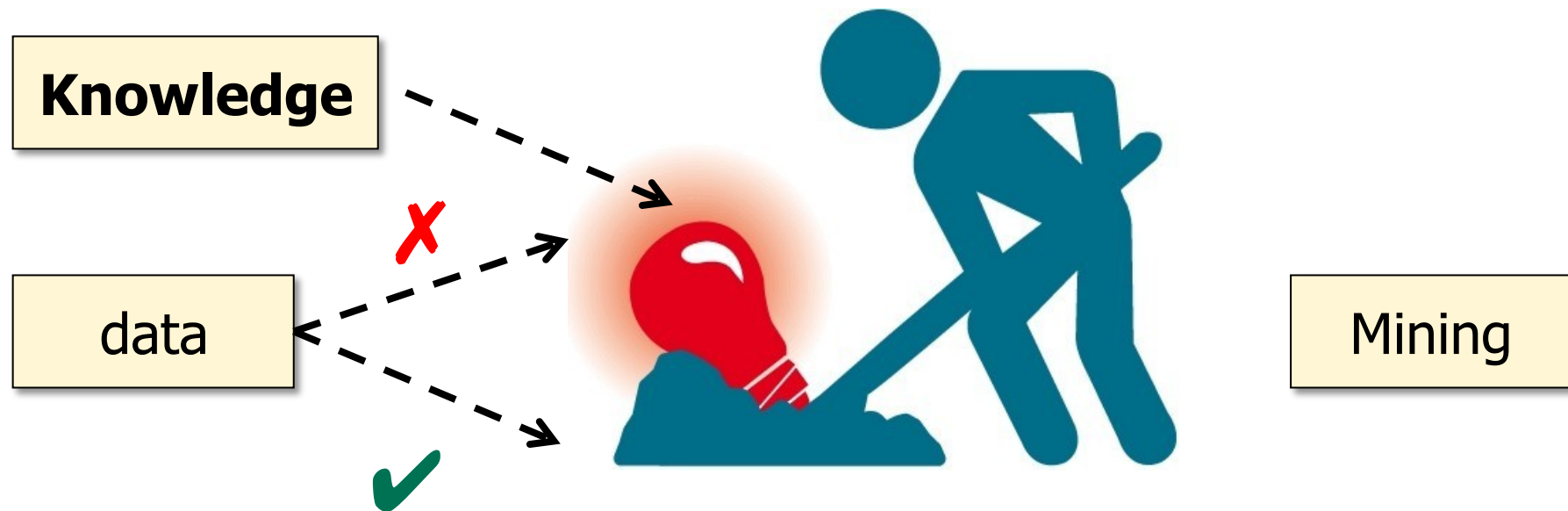
Example: Market Basket Analysis

- Data Mining @ Wal-Mart:
 - Some **obvious** correlations:
 - people who buy gin are also likely to buy tonic!
 - But one **interesting** correlation stood out
 - On Friday afternoons, **young American males** who buy diapers (nappies) also have a tendency to buy beer.
 - No one had predicted that result, so no one would ever have even asked the question in the first place.

What is data mining?



What is data mining?



We are drowning in **data**, but starving for **knowledge**!

What Is Data Mining?

- Alternative names
 - Knowledge discovery (mining) in databases (KDD)
 - knowledge extraction
 - data/pattern analysis
 - business intelligence
 - ...
- Data Mining (knowledge discovery from data)
 - Extraction of **interesting** (i.e., non-trivial, implicit, previously unknown and potentially useful) **patterns** or **knowledge** hidden in **large amounts of data**

Patterns



1. Association and Correlation
2. Classification
3. Cluster Analysis
4. Outlier Analysis

Next lectures...

Why Data Mining?

- *"Necessity is the mother of invention"*
 - Data mining—**Automated** analysis of massive data sets
- The Explosive Growth of Data: from terabytes to petabytes
 - Business: Web, e-commerce, transactions, stocks, ...
 - Science: Remote sensing, bioinformatics, scientific simulation, ...
 - Society and everyone: news, digital cameras, YouTube

Why Data Mining?



Patterns



1. Association and Correlation
2. Classification
3. Cluster Analysis
4. Outlier Analysis

Next lectures...