Data Mining
Intro Lecture

Jo Grundy

ECS Southampton

January 30, 2023

Data Mining - Introduction

Who are we?

- Dr Markus Brede (markus.brede@soton.ac.uk) 32 4033
- Dr Shoaib Ehsan (s.ehsan@soton.ac.uk) 32 4001
- Dr Jo Grundy (j.grundy@soton.ac.uk) 32 4053

Data Mining - Introduction

Who are you?

- MSc Students
- Fourth Year undergraduates

Data Mining - Introduction

Who are you?

- MSc Students
- Fourth Year undergraduates
- People who want to learn about data mining..

Vevox Poll 125-644-425

Data Mining - Introduction

Developed by Prof J. Hare
Between:

- Foundations of Machine Learning COMP3223
- Advanced Machine Learning COMP6208
- Machine Learning technologies COMP3222

Bridge between theory and practice

- How do you work with data?
- How do you solve real problems?

Data Mining - Introduction

Module Structure:

- Lectures:
  - Dr S. Ehsan: Regression, Information Theory
  - Dr J. Grundy: Data Mining Algorithms
  - Dr M. Brede: Networks
- Coursework 30%:
  - Group Project
- Exam 70%:
  - Computer Aided Multiple Choice/Calculation Questions.

Data Mining - Introduction

Lectures:

- Mondays 9 am
- Tuesday 9 am
- Thursday 12 pm
- Friday 3 pm

Generally we will only use the Monday, Tuesday and Thursday slots.
See the website for further details

Data Mining - Introduction

Coursework:

- Form a Group - 4-5
- Chose a topic
- Q & A sessions in week 4 will help with this
- Do a presentation before Easter
- Hand in write up at end of term

Data Mining - Introduction

Resources:

- website `http://comp6237.ecs.soton.ac.uk`
- Blackboard site

Reading:

- Data Mining and Machine Learning: Fundamental Concepts and Algorithms M. L. Zaki and W. Meira `https://dataminingbook.info/`
- Mining of Massive Datasets J. Leskovec *et al* `https://www.cambridge.org/core/books/mining-of-massive-datasets/C1B37BA2CBB8361B94FDD1C6F4E47922`

Data Mining - Introduction

What is Data Mining?

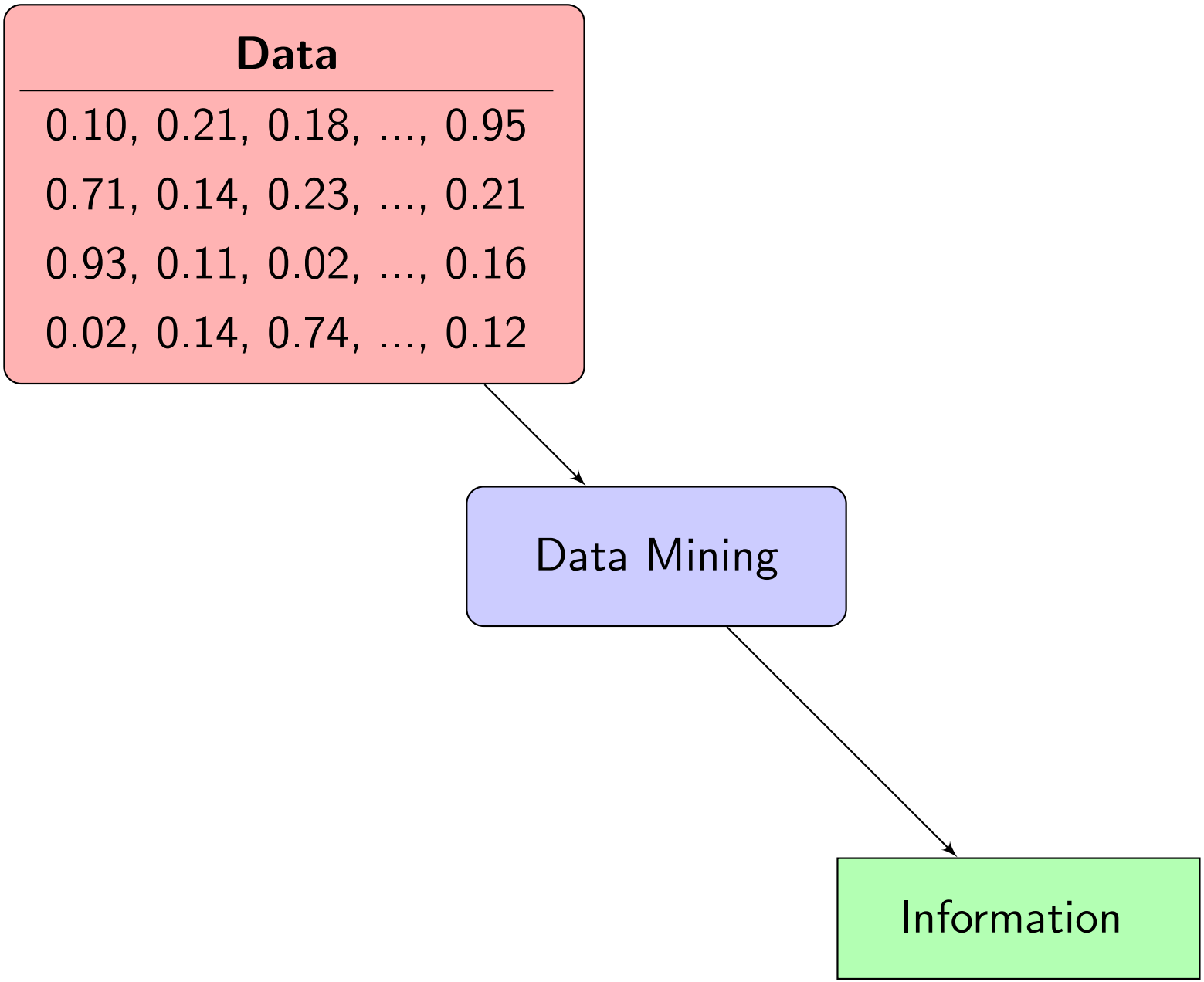
- Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. (wikipedia)

Data Mining - Introduction

What is Data Mining?

- Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Bill Palace

Data Mining - Introduction



Data Mining - Introduction

What is data?

- a sequence of numbers or symbols

It is not information, it needs interpretation.

Data Mining - Introduction

What is Information?

- "Actionable Knowledge" - C. Argyris
- Making predictions
- making sense

Data Mining - Introduction

So what is data mining?

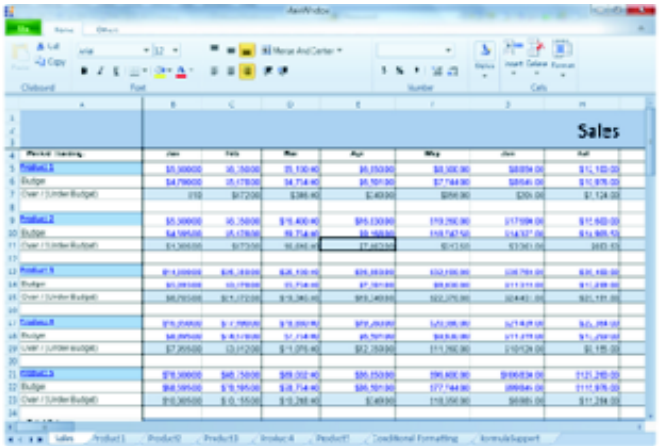
- Given lots of data:
- Discover patterns and models

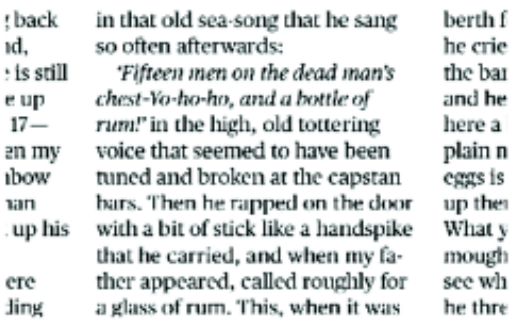
These patterns and models should be:

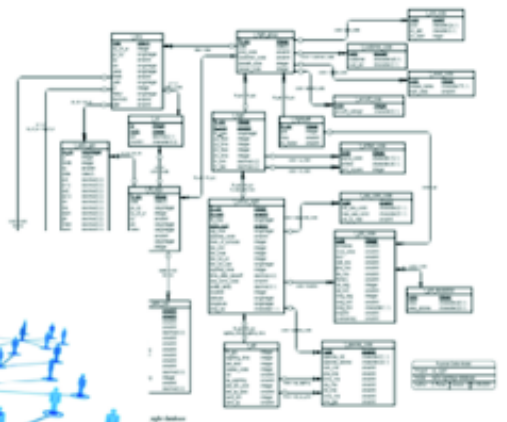
- Valid - hold for new data
- Useful - actionable
- Unexpected - not obvious
- Understandable - have human interpretability

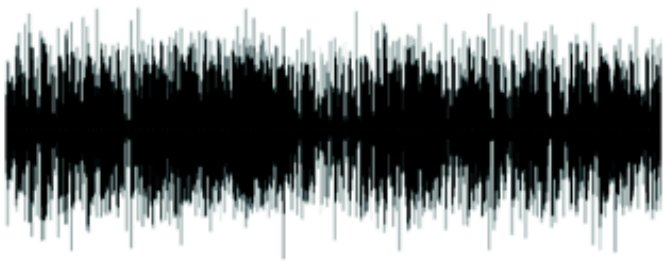
Data Mining - Introduction

What data can we mine?









Data Mining - Introduction

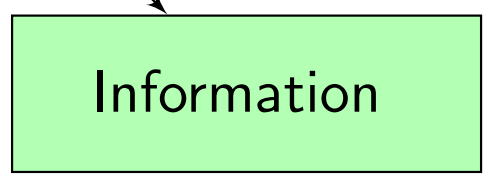
Data can be:

- Structured and Unstructured
- Dynamic/Static/Stream
- Unimodal/multimodal

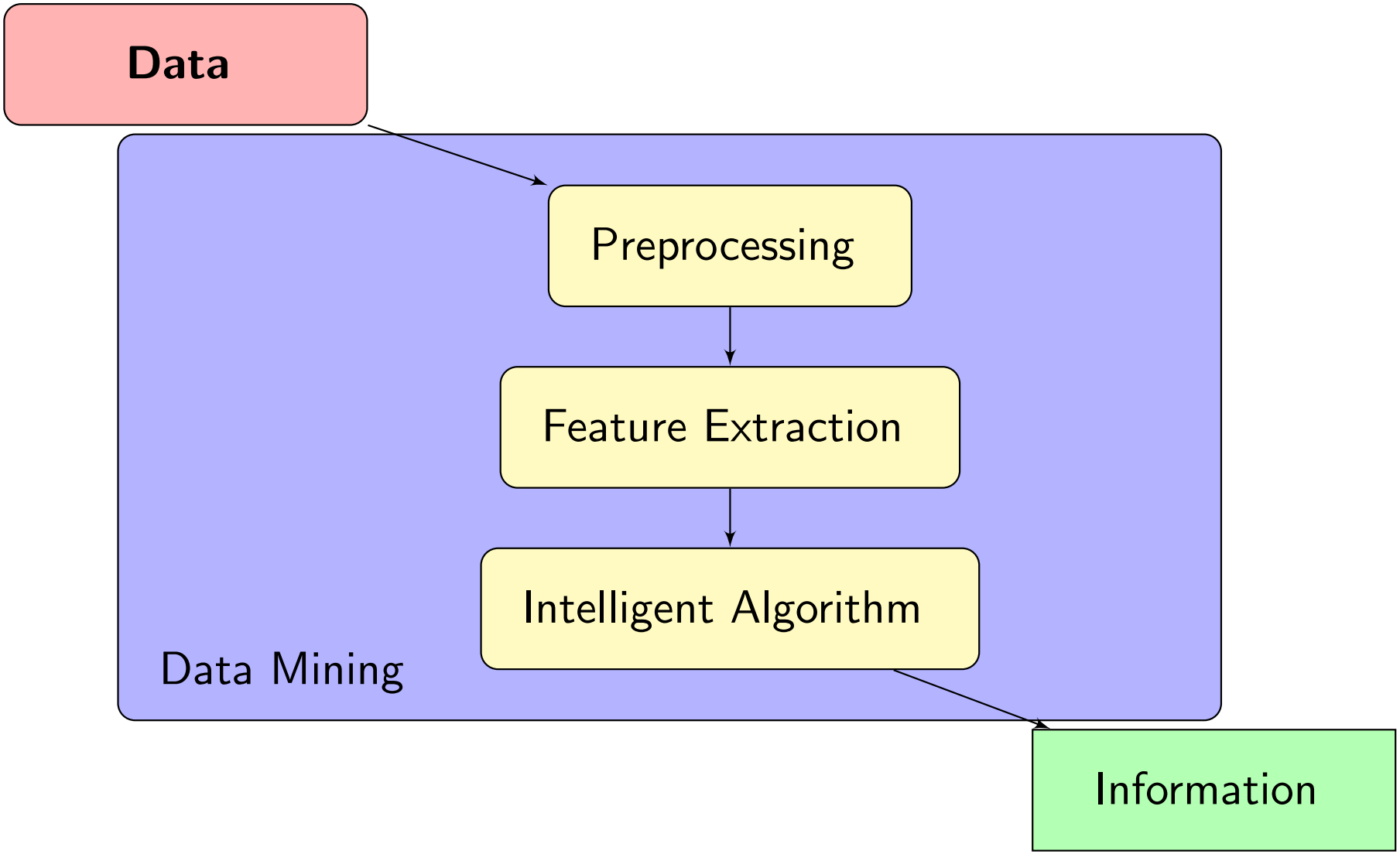
Data Mining - Introduction

Data

in that old sea-song that he sang
so often afterwards:
"Fifteen men on the dead man's
chest - yo ho ho and a bottle of



Data Mining - Introduction



Data Mining - Introduction

Intelligent Algorithms:

- Descriptive
  - PCA
  - Clustering
  - Anomaly Detection
  - ...
- Predictive
  - Classification
  - Ranking
  - Regression
  - Matrix Completion
  - ...

Data Mining - Introduction

The Plan:

- Shoaib
  - Regression
  - Information Theory
  - etc..
- Jo
  - Recommender Systems
  - Market Basket Analysis
  - Document Filtering
  - etc..
- Markus
  - Link Prediction
  - Community Detection
  - etc..

Data Mining - Introduction

Group Coursework:

1. Form Groups
2. Chose problem
3. Submit Brief - End of Week 4
4. Present ideas and approach to class - Week 8
5. Submit report - End of Term