



COMP60711 Part 2

Coursework Overview

# Goals of the Coursework

Developing your skills in discovering knowledge from data

Exposure to tasks you would do as a Data Scientist/Engineer

Not just about building models from data, but also reasoning about and explaining data through visualisation, analysis, and discussion

# Preparation Material

Make sure that you have followed the steps in the “Part 2 - Laboratory Preparation” files before starting the coursework

Covers:

- Setting up your coursework environment
- Using Jupyter notebooks
- Submitting your coursework
- A brief guide to plotting in Python

Any problems, come to the lab/drop-in sessions and ask for TA help

# Introductory Material

Opportunity to get setup and use to the tools used in the coursework

Covers:

- Basic usage of Python libraries: `scikit-learn`, `matplotlib`, and `seaborn`
- A graphical tool for data mining: Weka

If you have time, there are some additional tasks

We **strongly encourage** you to go through these exercises

## Please Note

MSc pass mark 50%

MSc pass with **Merit** 60%

MSc pass with **Distinction** 70%

# CW3 - Clustering & Itemset Rule Mining – Week 4

## Clustering

- 1) Clustering algorithmic behaviour and their sensitivity to data
- 2) Method for estimating the number of clusters
- 3) Applying clustering to a real-world dataset for knowledge discovery

## Itemset Rule Mining

- Alternative approach to classification/clustering/regression to find interesting relationships in data
- Applying a well-known method to real-world congressional voting records

**Deadline: Tuesday 24<sup>th</sup> Oct 10AM (Week 5)**

## CW4 – Classification / System issues - Week 5

- 1) Pre-processing & Feature Importance
- 2) Decision Boundaries
- 3) Training Time Comparison
- 4) Memory Usage Comparison

**Deadline: Tuesday 31<sup>st</sup> Oct 10AM (Week 6)**

## Key Points

**Make clear any assumptions and provide evidence to justify your answers**

**Cite sources; explain and justify your reasoning**

- If you need help, come to the online lab and/or drop-in sessions
- Submit as HTML
- Before submitting, check it!
- **Make sure it loads up properly and is error free**