1495 UNIVERSITY OF ABERDEEN | CELEBRATING 525 YEARS 1495 – 2020

ABERDEEN 2040

Introduction

JC3503: Data Mining & Visualisation

Lecture 1

2025

First off...

Welcome to JC3503: Data Mining and Visualisation!

In this course, you will learn the skills to analyse, visualise, and understand vast quantities of data.

We will discuss some of the approaches and methods for doing so, alongside some of the wider considerations within our data-driven society!

Today…

- Practicalities
- Assessments
- Schedule
- Learning outcomes
- About the module

Lecturer - AI

Dr. Chris Norval

Email: chris.norval@abdn.ac.uk

Office Hours: Thursdays 14:00 – 17:00

Research Areas:

Understanding Complex Data-Driven Systems (ML, IoT, Spatial Computing, etc.); Tech Law (data rights); Fair, Accountable and Transparent (FAccT) Systems; Human Computer Interaction (HCI)

Lecturer - BMIS

Dr. Jiangtian Nie

Email: jiangtian.nie@abdn.ac.uk

Office Hours: Thursdays 14:00 – 17:00

Research Areas:

Edge intelligence, wireless networking, and the Internet of Things.

Lecturer – CS

Dr. Siwei Liu

Email: siwei.liu@abdn.ac.uk

Office Hours: Thursdays 14:00 – 17:00

Research Areas:

I research on graph representation learning and its applications in recommender systems and AI4Science (Bioinformatics and Physics). In addition, I am passionate to develop fun tools using LLM.

Teaching Schedule

- Overall schedule:
  - 2 weeks of teaching (w/c 2025-04-14 – w/c 2025-04-21)
  - 1 week break (w/c 2025-04-28)
  - 2 weeks of teaching (w/c 2025-05-05 – w/c 2025-05-12)

- Each week of teaching
  - 8 lectures of course material
  - 4 practical labs (double sessions)

Assessments

- Assignment (25%)
  - Assignment will be released soon
  - Due after lectures finish

- Written Exam (75%)
  - Will cover all lectures
  - Calculator not permitted
  - Exact date and time TBC

Assignment

- Aim will be to analyse one or more pre-specified datasets using Python (and related data science libraries).
- Undertake exploratory analysis, find interesting attributes, patterns, and trends within the data, and report them.
- The goal is to demonstrate and apply the various techniques that we will cover throughout this course.

Learning Outcomes of JC3503

1. Manipulate, format, prepare, and clean data sets prior to analysis
2. Analyse complex datasets by applying data pre-processing, exploration, clustering and classification, time series analysis, and others
3. Design appropriate visualisation solutions for different applications, scenarios, and audiences

Technical Tools

- We will use Python and Jupyter this course:
  - Pandas for dataframes
  - Seaborn for visualisations (and datasets)
  - Scipy
  - Sklearn

- Other tools/packages are available!

Overall Schedule

- Week 1: Overview of Data Mining
  - Introduction to Data Mining
  - Exploratory Data Analysis (EDA)
  - Data Visualisation
  - A/B Testing and Null Hypothesis Statistical Testing (NHST)

Overall Schedule

- Week 2: Supervised Learning
  - Regression and Classification
  - Decision Trees
  - Naïve Bayes
  - Support Vector Machines

Overall Schedule

- Week 3: Unsupervised Learning
  - Clustering
  - Data Dimensionality
  - Association Rule Learning

Overall Schedule

- Week 4: Mining Different Types of Data
  - Time Series
  - Text Mining
  - Image Mining

Further Reading

Data Mining: Practical Machine Learning Tools and Techniques
Ian H. Witten & others (4th Edition, Morgan Kaufmann, 2017).

The Data Science Design Manual
Steven S. Skiena (Springer, 2017).

Data Visualisation: A Handbook for Data Driven Design
Andy Kirk (2nd edition, SAGE, 2019).

Data Mining & Visualisation

Broader Context & Background

ABERDEEN 2040

Data is Ubiquitous

In the modern world, data is everywhere...

… and vast quantities of data are being generated every day!

Social Data

Think about how much data a single person might generate:

- Messages
- Photos
- Location / GPS
- Online (and offline) purchasing
- Media consumption (music, films, websites visited)
- Social media content

Technical Data

Think about the data being generated by technical systems:

- Online services
- Banking systems
- Healthcare / medical devices
- Phones and computers
- Smart homes / IoT interactions
- Modern vehicles

Historic Data

Think about how much historic data is available!

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True

Data Analysis

Data is everywhere, and there is far too much for anyone to process and understand it themselves.

Modern computers give us the ability to perform sophisticated analyses on vast amounts of data, very quickly.

In this course, you will learn some of the skills to do this.

In short, aim is for you to learn to think like a data scientist!

Data Mining & Visualisation

Data Mining is the process of discovering patterns and extracting useful information from data.

Data Visualisation is the process of designing visual representations of data.

Fundamentally, both of these topics are about **understanding your data!**

Chat With Your Lecturer...

First off, what sorts of things do you want to learn?

How do you want these lectures to be structured? Conversational? Traditional?

Any questions about the course? About me?