OCMP5310: Principles of Data Science

Week 5 Live Session

Presented by

Daniela Rivas

THE UNIVERSITY OF
SYDNEY
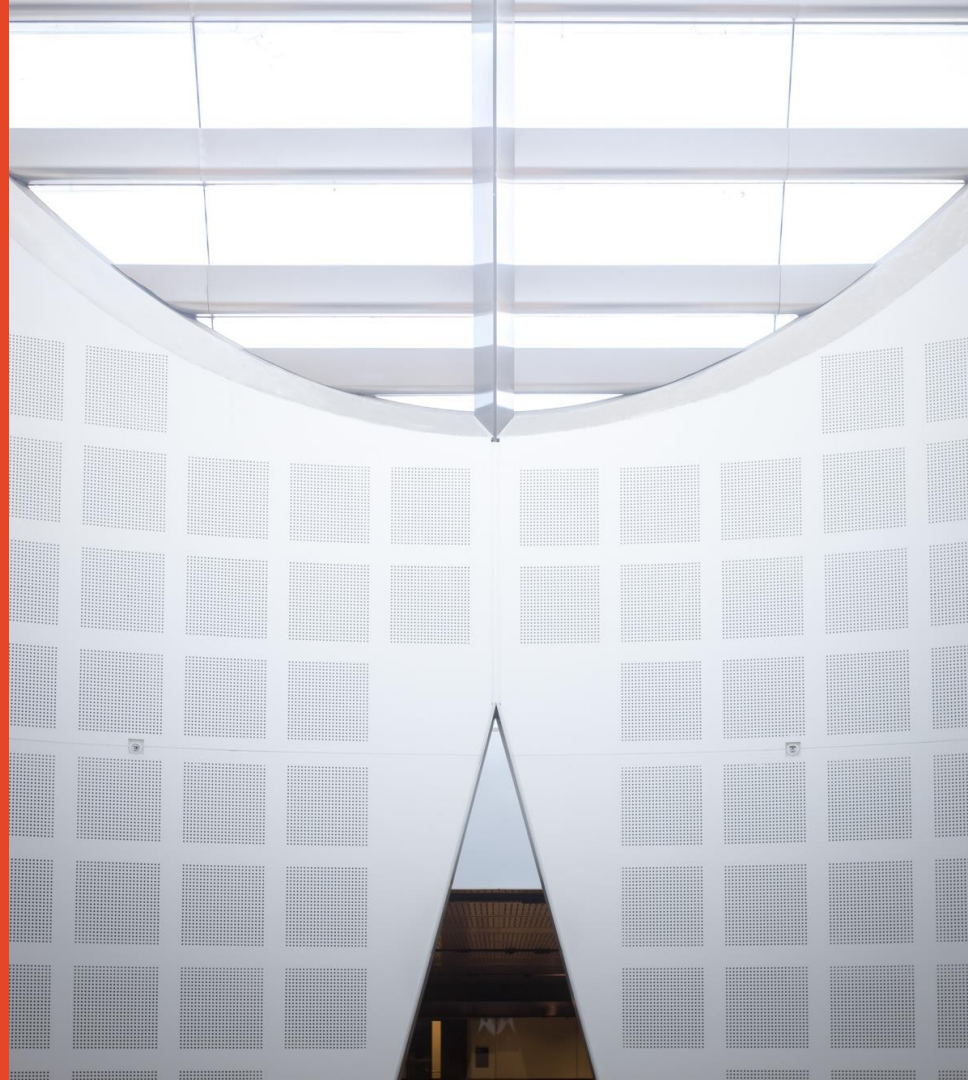
REVIEW OF LAST WEEK

Week 4

- – Unsupervised learning: Clustering.
- – Supervised learning: Regression and classification.

QUESTIONS?

WEEK 5: SESSION ACTIVITY

Week 5

- Text classification.
- General model evaluation.

Text Classification

Activity

- In Canvas, go to:
  - Exercise: Text Driven Forecasting.
- Download Jupyter Notebook:
  - exercises_unstructured_data.ipynb
- Download CSV files:
  - movies2013.csv
  - movies2014.csv

Model Evaluation

Activity

– In Canvas, go to:

    – Exercise: Model Evaluation.

– Download Jupyter Notebook:

    – from_data_to_decisions.ipynb

PROJECT STAGE 3 AND PRESENTATION

Project Stage 3

Project Stage 3 (15% overall mark)

- The objective of Stage 3 is to complete the analysis or model building to solve the problem you defined in Stage 1.

- **Due:** Week 6, 28 May 2023.

- **Tasks:**

  - Defining experiment setup.

  - Implementing your proposed approach.

  - Evaluating and analysing your approach.

- **Submissions:**

  - Report (4 pages max.).

  - Code.

Report

- No more than 4 pages.
- **Section 1: Setup**
  - State your research question(s).
  - Describe how you will quantify reliability. If appropriate, describe how you will measure effectiveness.
  - Identify datasets and the data you derived from them.
- **Section 2: Approach**
  - Describe your proposed model and any benchmark models for comparison.

Report

- **Section 3: Results**
  - Summarise results, reliability and comparison to any benchmark analyses or models.
  - Critically analyse results.
- **Section 4: Conclusion**
  - Conclude with what you have learned from this study which would improve yourself as a data analyst.

Code

– Your code used to build your predictive model.

– Format: Jupyter Notebook (.ipynb), python script (.py) or similar.

Presentation

Presentation (5% overall mark)

- Give a brief overview of your project, the data you investigate, and the current status of your work.
- **When?**
  - Week 6, 22 May 2023, during Live Session.

QUESTIONS?