# Group Project Stage 2 (Presentation)

Due: 11:59 PM on Sunday at the end of week 12 (Nov 2nd)

**Value: 8% of Total Mark**

Note: Get started your project ASAP. Discuss with your tutors and make use of Ed to ask questions.

## 1 Purpose

Stage 2 builds on the insights and collaboration developed in Stage 1. It introduces students to practical machine learning (ML) modelling and interpretation. Each group will design and compare different predictive models and deliver a group presentation during their Week 12 lab. Every member must present their own contribution.
This stage emphasises:

- Model development using Python
- Evaluation and interpretation of predictive models
- Communication of findings to an audience of peers with a data science background

## 2 Group Formation

- Groups must have 3–4 members and remain from Stage 1 unless changed with tutor approval.
- Any group may reuse, refine, or change their Stage 1 dataset.
- Everyone must contribute to the group coordination and shared presentation.

## 3 Project Work for Stage 2

This assignment should cover the followings:

## 3.1 Common Prediction Target

- Choose one attribute from a dataset to predict.
- The attribute can be nominal or quantitative.
- The dataset may come from Stage 1 or be newly selected.
- Preliminary analysis should suggest the target is meaningfully predictable.

## 3.2 Evaluation metrics

- Select at least one metric to evaluate prediction accuracy.
- Justify why this metric is appropriate.
- For higher marks, more than one metric should be used.

## 3.3 Data preparation

- Split the dataset into training, validation, and test sets (e.g. 75/15/10).
- No part of the test set may be used in model training or tuning.

## 3.4 Develop Predictive Models

- Use Python (e.g. scikit-learn) to train at least 3 families of predictive models on the training set. Note that at least 3 families of models must be used to answer the same research question/topic.
- Apply any necessary pre-processing, if needed.
- Tune hyperparameters using only the validation data.
- Present and explain modelling decisions.

## 3.5 Model Evaluation

- Evaluate each model's performance on the test set.
- Report at least one metric (more for higher marks)
- Interpret the performance and reflect on effectiveness or limitations.

### 3.6 Conclusion

- A comparison of the different approaches used
- Observations on which models worked better and why
- Any dataset-related insights or modelling challenges
- Reflections on metric choice and model limitations

### 3.7 Presentation

Each group will deliver one presentation (no longer than 10 minutes) during your assigned tutorial sessions in Week 12. Your presentation should be structured as follows:

- Your proposed modelling methods (note that at least 3 different predictive models must be used to answer your research question)
- Model evaluation - discuss different evaluation metric(s) used to assess model performance and provide interpretation of why/how they differ across different predictive models.
- Conclusion

You must present code, charts, or key outputs to support your findings/ explanation.

## 4 Stage 2 Submission

- **Submit on Canvas:**
  - **One PDF** containing the group's presentation slides.
  - The slide deck *must* include:
    - Slides for each model
    - The **group conclusion** slide(s).
- **Upload a compressed `.zip` archive** with the following structure:
  - **Predictive model subfolders** (one per model), each containing:
    1. Python code used to train and evaluate the model.
    2. Any cleaned datasets and/or transformation artifacts.

– **Shared content** folder (if applicable) for common resources.

Only one student per group needs to submit both files.
**Submission Links:**
**Presentation slides(.pdf) submission:** [Slides submission 1002]
**Code & Data submission:** `[Code & Data submission 1002]`

# 5 Marking

There are three components to be assessed. All are group-marked, and each member will receive the same score unless specific differences in contributions are explicitly noted.

| Component | Weight (%) | Full Marks Criteria |
|---|---|---|
| **Predictive modelling (max 6 slides)** | 40% | 1) Choose at least 3 different families of predictive models.<br>2) Clearly explain the model training process using Python (e.g. scikit-learn).<br>3) Show explicit code in the slides.<br>4) Demonstrate appropriate preprocessing and/or hyperparameter tuning.<br>5) Use a modelling technique distinct from those of other group members.<br>6) Present professionally (slides and verbal communication).<br>7) For full marks: justify why the chosen modelling approach suits the task. |
| **Model Evaluation (metric reporting + insight) (max 3 slides)** | 35% | 1) Compute at least one suitable evaluation metric per model on test data.<br>2) Clearly show the metric computation code.<br>3) Present professionally (slides and verbal communication).<br>4) For full marks: interpret results (e.g., overfitting/underfitting) and relate outcomes to dataset characteristics and model choices. Multiple metrics expected for higher marks. |

| | | |
|---|---|---|
| **Group Conclusion and delivery** | 25% | 1) Summary of findings in relation to your research topic<br>2) Clearly communicates relative strengths, weaknesses, and trade-offs of different models<br>3) For full marks: Presentation includes thoughtful discussion of modelling limitations, metric limitations, and reflects a shared understanding of machine learning principles<br>4) Each student must deliver at least one part of the presentation<br>5) Clear presentation (slide preparation and verbal communication) |

# 6 Late Submission

There is no late work of this assessment. If a student fails to turn up in Week 12 for the presentation, the student will receive zero mark for this assessment. However, if the Special Consideration of the student is approved for the absence of the class, the student will be given 3 relevant days to prepare a video to present his/her part of the work. If a student continues to apply for further Special Consideration and approved, the student will receive prorate a mark based on the student's Stage1 individual component.