Frequently Asked Questions (FAQ)

1. Is this a classification or regression task?

This coursework is a **classification** task.

Although the dataset contains a numerical *imdb_score*, you are required to convert it into 5 categories ("Poor" to "Excellent").

You do **not** perform numerical regression first.

2. Do we need to create the category labels ourselves?

Yes.

You must convert *imdb_score* into the 5 categories as specified in the brief.

3. After categorising IMDb scores, should imdb_score remain as a feature?

No.

Once you create the class labels, **remove** the original imdb_score from the predictors.

The categorical label becomes your target variable.

4. How do we handle actor names, director names, country, genre, etc.?

These are categorical variables.

You may use:

- Label encoding
- One-hot encoding
- Or dimensionality reduction for high-cardinality features

Justify whatever method you choose.

5. How many machine learning models must we use?

You must implement **at least three (3)** distinct supervised classification algorithms.

6. Can we include deep learning models like MLP?

No, classical ML algorithms are expected.

7. What train-test split should we use?

You may choose 70/30, 75/25, or 80/20, but you must explain the reasoning.

8. Is cross-validation compulsory?

Yes.

Use k-fold CV (k=5 or k=10 recommended) to evaluate model robustness.

9. What evaluation metrics should we use?

Use **classification metrics** as required:

- Accuracy
- Precision
- Recall
- F1-score
- Specificity

Justify your choice

10. Do we need confusion matrices or ROC curves?

Confusion matrices are strongly recommended.

ROC curves are optional for multi-class (one-vs-rest approach).

11. How do we compare the models?

Compare based on:

- Cross-validation performance
- Test set performance
- Strengths/limitations
- Metrics related to imbalanced categories.

Explain which model performs best and why.

12. How much hyperparameter tuning is required?

Tune only the **important** hyperparameters. Examples:

- SVM → C, kernel
- Random Forest → n_estimators, max_depth

you may extend as required for your work

Document:

- What you tuned
- Why you tuned it
- How it affected performance

13. What does "model interpretation" mean?

Interpretation means explaining **which features influence predictions**.

14. Should we include code in the report?

No.

The report must not include any code or screenshots of notebooks.

15. Can we include figures from the notebook?

Yes , you may include meaningful visual outputs from your notebook in the main report, such as:

- graphs
- tables
- charts
- evaluation plots (e.g., confusion matrix, ROC, feature importance)

These should **directly support your analysis** and must be integrated neatly in the *Results & Interpretation* section.

Important:

If you have additional figures, screenshots, or supporting materials that are useful but not essential for the main narrative, you should place them in the Appendix (A2), as stated in the coursework brief.

This keeps the main report clear and within the page limit, while still allowing you to include extra evidence of your work.

16. What are the page limits for the report?

- **Introduction** – 1 page
- **Methodology** – 4 pages
- **Results & Interpretation** – 2 pages
- **Summary** – 1 page
- **Appendix** – Maximum 4 pages

Strictly follow these limits.

17. How do we submit the Jupyter Notebook / Google Colab Notebook?

You **do not** upload the .ipynb file to Blackboard.

You must:

Step 1 : Convert your notebook to PDF

- From Jupyter: *File → Export As → PDF (or use any online converter)*
- From Colab: *File → Print → Save as PDF* (or File → Download → PDF)

Step 2 : Upload the PDF notebook to a cloud drive

Examples:

- OneDrive
- Google Drive

Step 3 : Share the link in the APPENDIX (A1)

In the Appendix section of your report:

- Paste a **single working link**
- Ensure permissions are set to **"Anyone with the link can view"**

📌 **Only the link should be included in the report, not the notebook pages.**

This is clearly requested in Appendix A1 of the brief.

18. Can we use GenAI tools for this coursework?

No.

This coursework is under **Tier 1 – No GenAI Use**.

This means:

- Do not paste your code into GenAI
- Do not ask GenAI to improve your report
- Do not upload any part of your assessment to AI tools

You **may** use GenAI **only to understand lecture slides or concepts**, not to generate assessed content.