

## Individual Assignment AM11

1. **Project Selection:** Choose a problem where you will use at least one out of the 5 topics that you have learnt to help solve a problem of your choice (CNN, SVM, Text Mining, PCA, Recommendation Systems).

- ⇒ The project should have a well-defined goal, such as classification, clustering, recommendation etc.
- ⇒ Plagiarism will result in 0 marks (e.g. replication of an existing Kaggle notebook). Your work must be original and well documented to explain your workings.
- ⇒ The complexity of your project should match the time available for submission.
- ⇒ The complexity of your work will reflect your grade (e.g. if you decide to work with a dataset that requires PCA pre-processing before classifying with SVM, thus utilising two out of five algorithms that you have learnt).

2. **Dataset:** Use an open dataset (e.g., from Kaggle, UCI ML Repository, etc.) or collect your own, ensuring it has enough samples but that it is not too large (you should be able to run your analysis on your laptop). For classification problems, ensure to properly balance your classes.

3. **Methodology:**

- Explain why the chosen technique is suitable for the problem.
- Preprocess the data appropriately.
- Train and evaluate the model using appropriate performance metrics.
- Compare with at least one baseline model

4. **Implementation (.py or .ipynb):**

- Use Python (with libraries like TensorFlow, Scikit-learn, Pandas, etc.).
- Ensure reproducibility (seed the random number generator where appropriate, provide a Jupyter Notebook (and its knitted output) or a well-documented .py script).

5. **Report (pdf):**

- Introduction: Explain the problem and dataset. Ensure to supply references. If you can produce your how to use TeX Studio and LaTeX.
- Methodology: Describe preprocessing, model selection, and training.
- Results & Discussion: Present evaluation metrics, visualizations, and insights.
- Conclusion: Summarize the findings and suggest future improvements.

Your report should be a maximum of 3 pages long, in an Arial 11 font with standard margins. Demonstrate the art of concise writing (brevity, economy of words, clarity and precision). Ensure your figure axes labelling and tickers are legible.

6. **Grading Criteria:**

You will be evaluated on both the technical execution and on your ability to communicate your findings.

Category	Weight	Description
Problem clarity & justification	20%	Clearly defines the problem, explains its relevance, and justifies the chosen ML technique.
Data preprocessing & exploratory analysis	20%	Properly cleans, preprocesses, and visualizes the data; identifies key patterns and challenges.
Model selection, training, and evaluation	30%	Implements an appropriate model, explains parameter choices, evaluates performance

		with meaningful metrics, and compares with a baseline.
Interpretation & discussion of results	20%	Provides insightful analysis, interprets results, discusses limitations, and suggests improvements.
Code quality & reproducibility	10%	Code is well-documented, structured, and reproducible; submission includes a Jupyter Notebook or well-commented script.