Department of Computer Science
Machine Learning and Neural Networks (COMP-5011)
Course Project Description Submission Guidelines

Total Marks: 100
Contribution to Final Assessment: 25%

1. Objectives

In this course project, you will serve the following objectives:

- Learn to extract the features from csv files or images (Dataset)
- Understand the role of feature extraction in machine learning
- Learn to use the different algorithms covered in the course to the problems of machine learning.
- Learn to document, report and present the solution of a machine learning problem
- Learn to evaluate and compare the performance of different algorithms

You are required to work in a **group of (maximum) three students** (every student in the group will receive the same score).

2. Kaggle Datasets

Kaggle is **the world's largest data science community** with powerful tools and resources to help you achieve your data science goals.

Students are required to use/implement machine learning techniques on their selected topic. It is essential for each student to contribute to this activity of learning. This will help you to understand the process of machine learning and how things are executing there.

Students are required to perform data cleansing, select appropriate classification methods and execute their selected method on the provided dataset and measure the performance at the end.

<u>If your work will have a worth to publish then we may proceed for the publication too.</u>

3. Expectations and Scope of Work

3.1 Scope of Work

The scope of the project includes,

Task 1: Formulation of the problem under consideration.

Task 2: Carry out literature review to identify the features that can be used for the problem under consideration.

Task 3: Apply feature extraction/engineering

Task 4: Apply dimensionality reduction (if needed or to evaluate the impact of reduction on the performance).

Task 5: Implement the different classifiers for the problem under consideration.

Task 6: Report the performance of different classifiers and presentation of analysis/findings.

3.2 Assessment

There are three components of assessment in the project.

- Project Report (20 marks)
- Project Code (15 marks)
- 8-10 minutes presentation (10 marks)
- Deliverable Submissions on time (5 marks)

For a more specific description, consult Project Rubrics.

Final report must be prepared using the provided template (to be uploaded on D2L).
Your report is expected to have the following sections.

- Abstract (Executive Summary)
- Introduction
- Mathematical Formulation
- Identification and Extraction of Features
- Feature engineering, dimensionality reduction (optional)
- Performance Evaluation (plots or tables), Analysis and Findings
- Conclusion

Note: We encourage students to use latex templates for their final report.

4. Timeline of Deliverables

Students are required to adhere to the following timelines.

Deliverable 1: Project Selection and Group Formation

- Due: Week 4, Thursday September 25, 2025
- Only one group member will send an email to the instructor to inform about group members and the project topic.

Deliverable 2: Submit a preliminary project report with the following sections populated. You will add contents in these sections as per your final submission.

- Abstract, Introduction and Data Analysis

  Due: Week 8, Thursday October 23, 2025 @ 23:55

Deliverable 3: Submit a report and code with the following tasks completed and added in the report

- Implementation of feature extraction
- Feature Engineering (e.g., dimensionality reduction)
- Implementation of selected classification algorithm (Supervised Learning)

  Due: Week 10: Thursday November 06, 2025 @ 23:55

Deliverable 4: Submit code (documented), final report and presentation (8-10 minutes)

Due: Week 12: Final project presentations are scheduled on Monday and Wednesday (24 and 26 November 2025 during the lecture time)

The code and final report with presentation slides must be uploaded via a link on D2L on November 23, 2025.

(No extensions will be given)

Project Rubric

Criteria	Excellent (5) 88-100	Good (4) 78-87	Satisfactory 70-77	Needs Improvement 60-70	Unsatisfactory (1) <60
Problem Definition and Understanding **(15)**	The problem is well-defined, and the student demonstrates a deep understanding of its significance and context. Clear problem statements, objectives, and relevant domain knowledge are provided.	The problem is adequately defined, and the student shows a good understanding of its significance and context. Problem statements and objectives are clear, but some additional domain knowledge could be included.	The problem is defined, but there might be some ambiguity. The student shows a basic understanding of the problem's context. Problem statements and objectives lack clarity or context.	The problem definition is vague or unclear. The student struggles to convey the problem's significance and context. Problem statements and objectives are poorly defined.	The problem is not defined, or it is completely unrelated to the course content.
Data Collection and Preprocessing **(10)**	The student demonstrates exceptional data collection skills, using diverse and relevant sources. Data preprocessing is thorough and effectively handles missing values, outliers, and noise.	The student collects suitable data from multiple sources. Preprocessing is well-done, addressing most missing values, outliers, and noise.	The student collects some relevant data, but there might be limitations in the sources or diversity. Preprocessing is basic, with some issues in handling missing values, outliers, or noise.	The data collection lacks relevance, or the student relies on a single source. Preprocessing is minimal, with significant issues in handling missing values, outliers, or noise.	Data collection is absent or entirely inappropriate. Preprocessing is missing or severely flawed.
Feature Engineering and Selection (15)	The student demonstrates exceptional feature engineering skills, creating highly relevant	The student engineers relevant features, and some feature selection techniques are	The student performs basic feature engineering, with limited exploration of feature selection	Feature engineering is minimal, and the student struggles to implement feature selection techniques.	No evidence of feature engineering or selection.

	and informative features. Feature selection methods are advanced and effectively applied.	applied. Features are generally useful, but there might be room for improvement.	methods. Features are somewhat relevant, but there is a lack of creativity.	Features are rudimentary and lack relevance.	
Model Selection and Training **(20)**	The student selects appropriate models based on in-depth analysis and understanding of the problem. Models are trained with extensive hyperparameter tuning, and ensemble methods might be used effectively.	The student selects suitable models, considering the problem's characteristics. Models are trained with reasonable hyperparameter choices. Some basic ensemble methods might be applied.	The student's model selection is somewhat limited in scope or relevance. Basic hyperparameter tuning is performed. Ensemble methods are attempted but not well-integrated.	The student struggles to select appropriate models and set hyperparameters . Limited or no use of ensemble methods is evident.	Inappropriate model selection, no hyperparameter tuning, and no use of ensemble methods.
Model Evaluation and Performance **(15)**	The student demonstrates a comprehensive evaluation strategy, employing a variety of relevant metrics. The model's performance is outstanding, showing high generalization to unseen data.	The student conducts thorough model evaluation, using appropriate metrics. The model's performance is good and demonstrates reasonable generalization.	The student's model evaluation is basic, and there might be some mismatch between chosen metrics and the problem. Model performance shows room for improvement in generalization.	Model evaluation is minimal, and the student struggles to interpret the results. The model's generalization is poor.	Model evaluation is absent or entirely inappropriate.
Interpretation of Results and Analysis **(10)**	The student provides exceptional insights into model behavior, potential biases, and failure modes.	The student offers meaningful insights into model behavior and potential biases. Some basic analysis techniques are applied.	The student's interpretation is limited to surface-level observations. Analysis techniques, if present, lack depth.	The student struggles to interpret results, and analysis is minimal or unclear.	No interpretation or analysis provided.
Communication and Presentation	The project report/presentation is	The project report/presentation is	The project report/presentation is adequately	The project report/presentation lacks	Poorly organized project report/presentation

(5)	exceptionally well-organized, clear, and engaging. Visualizations are highly effective in conveying information.	well-organized and clear. Visualizations are effective.	organized and clear, but there might be room for improvement. Visualizations are somewhat effective.	organization or clarity. Visualizations are confusing or irrelevant.	with unclear content and inadequate visualizations.
Code Quality and Documentation **(5)**	The code is exceptionally well-structured, well-documented, and follows best practices. Comments and documentation are thorough, facilitating easy understanding and replication.	The code is well-structured and adequately documented. Best practices are followed, and comments aid in understanding.	The code structure is reasonable, but some improvements could be made. Documentation is present but might lack details in some areas.	Code structure is disorganized or unclear. Documentation is sparse or lacking crucial information.	Code is poorly structured, lacking documentation and adherence to best practices.
Collaboration and Contribution **(5)**	The student actively engages with peers, seeks and offers help, and contributes significantly to group discussions.	The student collaborates effectively with peers, participates in discussions, and contributes positively.	The student participates but might lack active collaboration or substantial contribution to group dynamics.	The student's collaboration is limited, and contributions to group activities are minimal.	The student does not participate in collaboration or contribute to group efforts.