Principles of Data Mining and Machine Learning (MOD 007892, Element 010, Component 1)

As part of the assessment for Element 010, Component 1, you are expected to submit the following three files through a single Canvas link:

- Jupyter notebook of your coursework project
- Coursework project report
- Lab Logbook (logbook must contain GITHUB or sharable one drive link of all your source code)

Component 1, Part 1 – Coursework Project (60% of the total module marks)

In this assessment, you will implement and evaluate multiple machine learning models on a real-world ECG Heartbeat dataset and provide a report based on your experiments and analysis.

Electrocardiogram (ECG) can reliably be used to monitor the functionality of the cardiovascular system. Recently, there has been significant attention towards the accurate categorization of heartbeats. In this project, the objective is to employ a Machine Learning approach for the classification of ECG signals. The link to the dataset has been provided below:

https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data

The link provides access to two different datasets: the Arrhythmia Dataset and the PTB Diagnostic ECG Database. Out of these two datasets, you are required to use the Arrhythmia Dataset for this project. In this assessment, your task is to develop machine learning-based classifiers to categorize Arrhythmia signals. You are encouraged to be creative and have the freedom to choose any machine learning algorithm, even if it hasn't been covered in the classroom. The assessment includes the submission of a Jupyter notebook and a written report.

Jupyter Notebook:

1. You must write a working code for this task. The proposed steps for writing your code are as follows:
2. Download the dataset from the provided link.

3. Perform exploratory data analysis and visualization on the dataset.
4. Pre-process the data to identify the best features for acceptable performance.
5. Develop supervised or unsupervised Machine Learning algorithms and compare their performances.
6. Implement any novel solution for the problem addressed.
7. Compare the performance of the different algorithms in terms of accuracy and other metrics.

Written Report:

You need to write a report based on the experiments. Your report should be divided into the following sections.

1. **Introduction:** Introduce the problem and explain why you are using ML as your solution (maximum 1-2 paragraphs).
2. **Exploratory Data Analysis and Data Visualization:** Display the key characteristics of the dataset and provide a summary in the form of graphs.
3. **Implementation:** Explain the ML pipeline and the novel solution you have adopted. You need to explain the insights of the data processing steps as well as the ML methods adopted, explaining them in terms of mathematical formulations.
4. **Results:** Show and explain the results of your evaluation. Commonly used metrics are accuracy, F-1 score, Precision, Recall, and the confusion matrix. However, you are free to choose other metrics if you prefer.
5. **Conclusions:** Explain what needs to be done to further improve the performance of the classifier (maximum 1 paragraph).
6. **References:** Cite in the text and add any references.

Note: You do not need to include snapshots of the code or explain it line-by-line in your report. Any code explanation will not be marked.

Submission requirements for this part: You need to submit both the written report and code through Canvas as separate files.

Component 2, Part 2 – Lab Logbook Submission (10% of the total module marks)

For each tutorial session, we will request specific details or summaries of the work you have conducted. For example, you may be required to include graphs or charts in your lab logbook for certain experiments.

You are expected to include the specified summaries for 10 tutorial sessions, as outlined on Canvas. To ensure you have the necessary information for each lab session, please refer to the Canvas page for each week, where you will find the specific instructions and details to be added in this section. Add the specific details or summaries in the relevant section of this document.

It's important to note that you should incorporate the specific details or summaries for each tutorial session into this singular logbook. There's no need to create separate logbooks for each lab; instead, you will maintain one logbook where you consistently record the details for each week's activities.

Submission Requirement for this part: In this part, you are required to submit a single lab logbook that you have maintained throughout the semester, including the specified summaries for 10 tutorial sessions. Within your lab logbook, provide the link to your GITHUB or a shareable OneDrive folder containing all your source code. The evaluation will primarily be based on your lab logbook, but we will also review your code to confirm that the logbook accurately represents the output from your code.

Rubrics

Description	Marks
Introduction	5
EDA and data visualization	10
Implementation (ML algorithms)	20
Results	15
Conclusion	5

Code quality	5
Lab logbook	10