DEPARTMENT OF COMPUTER & INFORMATION SYSTEMS ENGINEERING
BACHELORS IN COMPUTER SYSTEMS ENGINEERING

Course Code and Title: CS-324 Machine Learning
Complex Engineering Problem
TE Batch 2022, Spring Semester 2025

TERM PROJECT

Course Learning Outcome

CLO 2: Investigate techniques and principles for implementing machine learning systems. (Taxonomy level C4, PLO-4 Investigation).

Complex problem-solving attributes (CPA) covered (as per PEC - OBA manual – 2019)

- **CPA-1 Depth of analysis required:** Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models.
- **CPA-2 Level of interaction:** Require resolution of significant problems arising from interactions between wide ranging or conflicting technical, engineering or other issues.
- **CPA-3 Familiarity:** Can extend beyond previous experiences by applying principles-based approaches.

Description

You are required to develop a machine learning system to classify the output as indicated below for any one of the following datasets.

Dataset title	Type of data	Label to be predicted	Download link
Weather Dataset	Structured tables	Multiclass classification – Predict the 'Summary' attribute. You may reduce the number of classes to up to 3 for better performance.	https://www.kaggle.com/datasets/muthuj7/weather-dataset
CIFAR-10	Images	Binary classification – Create a new label by relabeling the outputs into two classes, namely 'animal' and 'vehicle'.	https://www.cs.toronto.edu/~kriz/cifar.html

Your assignment should encompass the following requirements:

- Perform all necessary data preprocessing steps on the dataset.
- Implement at least three algorithm: one non-parametric algorithm, one parametric shallow algorithm, one neural network architecture.
- For each algorithm provide at least 3 models by selecting different values of hyperparameters.
- Ensure proper development and test validations to select the best model.
  Also, develop a simple interface (within the notebook) to allow the user to test your system by providing input and receiving the predicted label in return.
- Report your comparative results for all nine models in both tabular and graphical forms.
- Investigate the effect of changing the data split ratio on the three models selected for each algorithm.

Some Extension Ideas

- Try out more hyperparameters than minimally required.
- Try out more algorithms than minimally required.
- Try out incremental learning. Your interface should provide the user with an additional option of incremental training. So if user provides a new dataset (similar in structure to the one on which your algorithm was trained previously), then your algorithm should be able to incorporate the new knowledge. Your code should be able to perform all required pre-processing steps too on this new dataset.
- Try out some transfer learning techniques.

Instructions and Guidelines

- Students can work in groups of 3 at maximum.
- Your assignment will be graded on the attached rubric (see Google Classroom).
- The code must be easy to read and follow. Provide concise and useful comments/markdown text generously.
- You may add any additional features to the project to make it distinct and stand out.
- You may refer to any web resources or books. Do refer the resource(s) in your report.
- The project shall be graded for 20 marks in final practical exam, where you will demonstrate working of the project and answer relevant questions.

Deliverables

- Code on Jupyter notebook
  - Develop the code in a single Jupyter notebook.
  - Add a markdown cell at the top showing roll numbers and names of all group members.
  - Save the notebook with the following naming convention: CS22XXX, where XXX represents roll number of any one group member in 3 digits. For example: CS22032.py
  - Add comments generously.
- Prepare a report organized as follows:
  - Details of the data preprocessing steps applied.
  - Details of the models and machine algorithm chosen for implementation.
  - Details of any distinguishing features (if any).
  - Tabular and graphical comparison of the results of all the models.
  - Comments on the performance of the implemented machine learning system, including issues like underfitting and overfitting, suggesting any techniques for improvement.
  - Use font size of 12 and/or 14 for headings, and 11 for regular text. Font style should be Times New Roman.
  - Attach the provided rubric on **top** of your report, with names and roll numbers filled in. This will serve as the title page of your report; no need to attach any other title page.
  - Save the report as pdf file, with the following naming convention: CS22XXX, where XXX represents roll number of any one group member in 3 digits. For example: CS22032.py.

Submission

- Online submission:
  - Submit both the files, notebook (ipynb file) and report (pdf file) with the specified naming convention on google classroom latest by **May 19, 2025**.
  - Any one of the group members can do the submission.
- Bring the notebook (softcopy) and report (hardcopy) during final practical exam in **Week 15**.