CSE381 Introduction to Machine Learning – Fall 2023

Project Description

Using Titanic Dataset from Kaggle platform https://www.kaggle.com/competitions/titanic , it is required to design different classifiers. For the training set, it is provided the outcome (also known as the "ground truth") for each passenger. Your model will be based on "features" like passengers' gender and class. You can also use feature engineering to create new features.

You will have to divide the data into three partitions: training, validation, and testing. You need to design the following classifiers:

1. Support Vector Machines
2. K-Nearest Neighbors
3. Bayes Classifier (**due end of week 11**)
4. Decision Trees
5. Multi-Layer Perceptron (**due end of week 13**)

For each classifier, try to find the optimal hyperparameters. You also need to compare the performance of all classifiers using different metrics such as the precision, recall, F1-score, and ROC/AUC curves.

Deliverables:
- Detailed report containing steps, screenshots of the code, screenshots of the output, visualization of the accuracy change, the outputs and show the reason of using the final values of the hyperparameters.
- Python code of the whole project
- A readme file containing the steps to run the project and what libraries are to be imported if any.
- You can make a group up to 3 students.
- Submission should be in LMS

Marks Distribution:

SVM	**15%**
KNN	**15%**
Bayes classifier	**20%**
Decision Tree	**15%**
Multi-Layer Perceptron	**15%**
Documentation	**20%**