Project– Spring 2024

Course Code	Course
CSE 381	Introduction to Machine Learning
Instructor/s	Dr. Mahmoud Khalil, Eng. Mahmoud Soheil

Project Description

This project is developed in teams of 2-3 students.

In this project, you will design five classifiers: a naive Bayes classifier, a SVM classifier, a KNN classifier, MLP classifier and a decision tree. You will test your classifiers on the 3 data sets.

First dataset in breast_cancer_diagnosis.csv is breast cancer diagnosis which contains features which are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. There are two classes in Diagnosis column which are (M = malignant, B = benign).

Second dataset is about the past loans. The Loan_train.csv data set includes details of 346 customers whose loan are already paid off or defaulted. The features are described as below while loan status is the output:

Field	Description
Loan_status	Whether a loan is paid off on in collection
Principal	Basic principal loan amount at the
Terms	Origination terms which can be weekly (7 days), biweekly, and monthly payoff schedule
Effective_date	When the loan got originated and took effects
Due_date	Since it's one-time payoff schedule, each loan has one single due date
Age	Age of applicant
Education	Education of applicant
Gender	The gender of applicant

Third dataset in Iphone_purchase.csv which tells which of the users purchased/not purchased iPhone. The features are gender, age and salary

Project Milestones

The two milestones shall be submitted by one of the team on the LMS and mention the team members with their IDs on cover page. The deadlines are shown on LMS.

➢ First Milestone

Requirements:
- Load the dataset
- Train the Naïve Bayes Classifier and KNN classifier on dataset and change the hyperparameters (if exist)
- Test the classifiers on the testing dataset and show the accuracy for each run when changing the hyperparameters

Deliverables:
- Detailed report containing steps, screenshots of the code, screenshots of the output, visualization of the accuracy change, the outputs (faces detected or numbers detected) 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 to be imported if any.

➢ Second Milestone

Requirements:
- Train the MLP Classifier, SVM classifier and build decision tree on both datasets and change the hyperparameters (if exist)
- Test the classifiers on the testing dataset and show the accuracy for each run when changing the hyperparameters

Deliverables:
- Detailed report containing steps, screenshots of the code, screenshots of the output, visualization of the accuracy change and the outputs (faces detected or numbers detected) 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 to be imported if any.

Marks Distribution:

Loading of dataset	**10%**
Naiive bayes classifier	**10%**
KNN	**10%**
MLP	**20%**
SVM	**15%**
Decision Tree	**15%**
Documentation	**20%**