ITD105 Case Study #1

Comparing Machine Learning Algorithms

Name: _____

I CLASSIFICATION

Train the **classification dataset** using various machine learning algorithms designed for classification. Evaluate and compare these models by applying different resampling techniques and utilizing appropriate performance metrics.

Classification Dataset

Dataset Name : _____

Features: _____

Set A
Resampling Technique : _____
Classification Metric : Confusion Matrix and Classification Report

ML Algorithm (Classification)	**Confusion Matrix** ***(Provide the matrix and classification report of each algorithm)***
CART (Classification and Regression Trees)	
Gaussian Naive Bayes/Naive Bayes	
Gradient Boosting Machines (AdaBoost)	
K-Nearest Neighbors (K-NN)	
Logistic Regression	
Multi-Layer Perceptron (MLP)	
Perceptron	
Random Forest	
Support Vector Machines (SVM)	

Set B *(should use different resampling technique and classification metric)*
Resampling Technique: _____
Classification Metric: _____

ML Algorithm (Classification)	
CART (Classification and Regression Trees)	
Gaussian Naive Bayes/Naive Bayes	
Gradient Boosting Machines (AdaBoost)	
K-Nearest Neighbors (K-NN)	
Logistic Regression	
Multi-Layer Perceptron (MLP)	
Perceptron	
Random Forest	
Support Vector Machines (SVM)	

Set C *(should use different resampling technique and classification metric)*
Resampling Technique: _____
Classification Metric: _____

ML Algorithm (Classification)	
CART (Classification and Regression Trees)	
Gaussian Naive Bayes/Naive Bayes	
Gradient Boosting Machines (AdaBoost)	
K-Nearest Neighbors (K-NN)	
Logistic Regression	
Multi-Layer Perceptron (MLP)	
Perceptron	
Random Forest	
Support Vector Machines (SVM)	

Results interpretation (Set A , Set B and Set C):

Based on the results, perform algorithm/hyperparameter tuning (at least 3) of the chosen ML algorithm.

EXAMPLE:

ML Algorithm: Support Vector Machines (SVM)
Sampling Technique - Train/Test Split (80:20)
Classification Metrics – Accuracy

	SVM Hyperparameters			
	random_state	**Kernel**	**C**	**Accuracy**
Model I	seed	linear	1.0	77.922
Model II	5	poly	1.5	79.870
Model III	10	poly	2	79.221

Results interpretation:

II REGRESSION

Train the **regression dataset** using various machine learning algorithms designed for regression. Evaluate and compare these models by applying different resampling techniques and utilizing appropriate performance metrics.

Regression Dataset
Dataset Name : _____
Features: _____

Set A
Resampling Technique : _____
Regression Metric : _____

ML Algorithm (Regression)	
CART (Classification and Regression Trees)	
Elastic Net	
Gradient Boosting Machines (AdaBoost)	
K-Nearest Neighbors (K-NN)	
Lasso Regression	
Ridge Regression	
Linear Regression	
Multi-Layer Perceptron (MLP)	
Random Forest	

Set B *(should use different resampling technique and regression metric)*

Resampling Technique: _____

Regression Metric: _____

ML Algorithm (Regression)	
CART (Classification and Regression Trees)	
Elastic Net	
Gradient Boosting Machines (AdaBoost)	
K-Nearest Neighbors (K-NN)	
Lasso Regression	
Ridge Regression	
Linear Regression	
Multi-Layer Perceptron (MLP)	
Random Forest	

Results interpretation (Set A and Set B):

Based on the results, perform at algorithm tuning (at least 3) of the chosen ML algorithm.

EXAMPLE:

ML Algorithm: Support Vector Machines (SVM)
Sampling Technique - Train/Test Split (80:20)
Regression Metrics – MAE

	SVM Hyperparameters			
	epsilon	**Kernel**	**C**	**MAE**
Model I	0.1	rbf	1.0	5.754
Model II	0.2	linear	1.5	3.754
Model III	0.15	poly	1.25	5.761

Results interpretation:

Submit the following:

a. Pdf copy of the results.
b. Video link demonstrating the case study.