Machine Learning - Assignment 1

Submitted By: Himanshu Aggarwal
Roll No.: MT17015

Prediction accuracies for the given datasets using the following models:

	Logistic Regression	Gaussian NB	Decision Tree Classifier
Dataset Part A	0.845	0.579	0.729
Dataset Part B	0.579	0.571	0.607
Dataset Part C	0.996	0.958	0.946

Given in the table are the mean accuracies respective to the model, after implementing grid search and K-Fold cross validation for K=3.

Details:

Logistic Regression: This model is implemented with grid search iterating on hyperparameter *C*. *C* is the inverse of the regularisation strength ($C = 1 / lambda$). All graphs of this models are values of C v/s Accuracy of model.

Gaussian NB: This model is implemented without grid search since the hyperparameter *prior,* which is the probability of the classes, is calculated according to the given data. Hence, no parameter is there to test the accuracy on.

Decision Tree Classifier: This model is implemented with grid search iterating on hyperparameters *max_depth* and *min_samples_split.* All graphs of this models are values of hyperparameter v/s Accuracy of model.
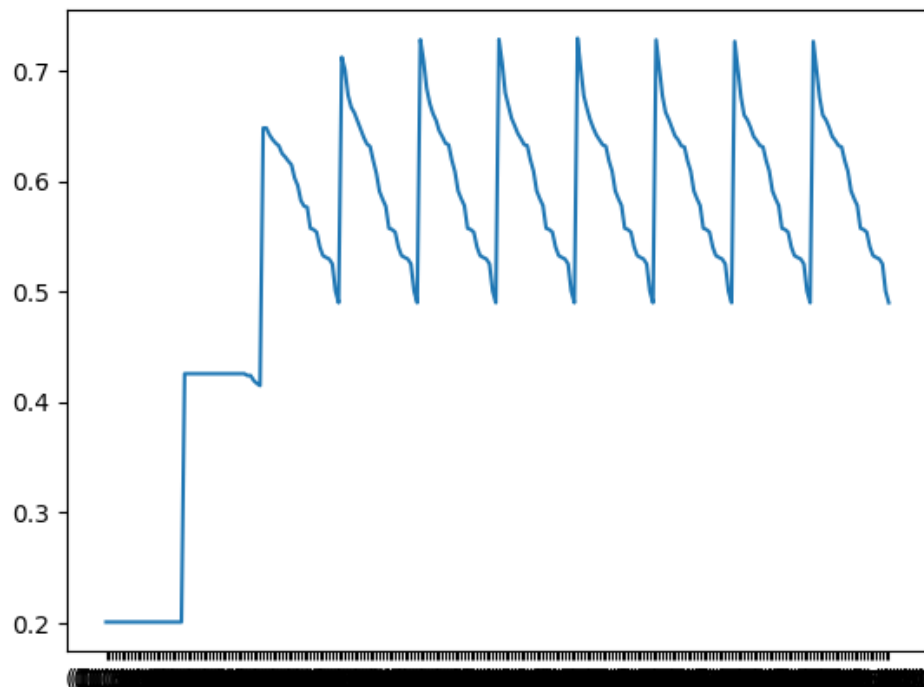
Dataset Part A

This is the visualisation for this dataset. This clearly shows clustering of the 10 different classes in different colours.
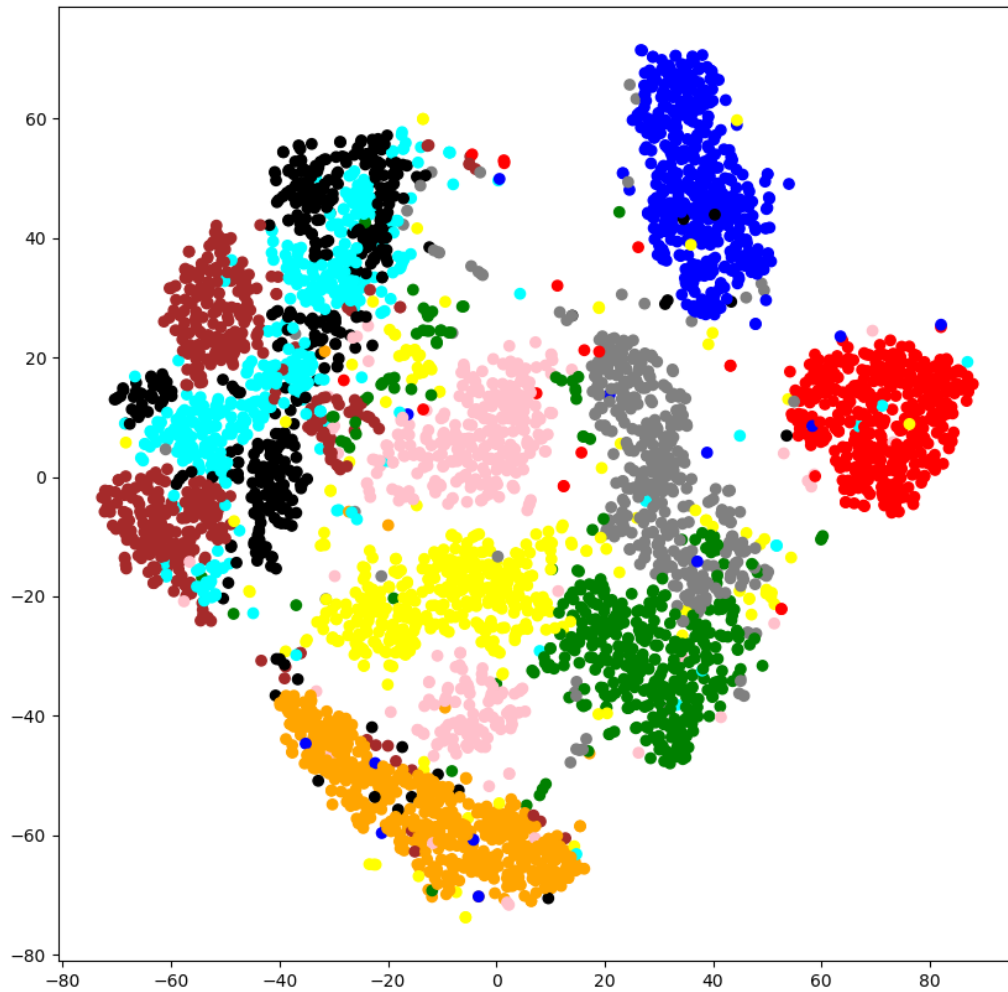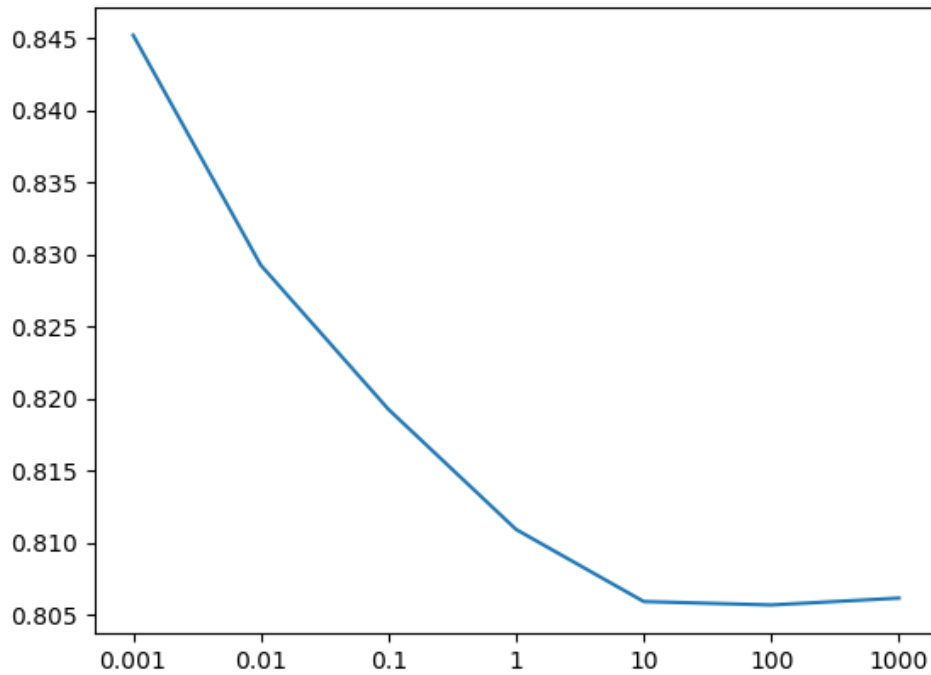
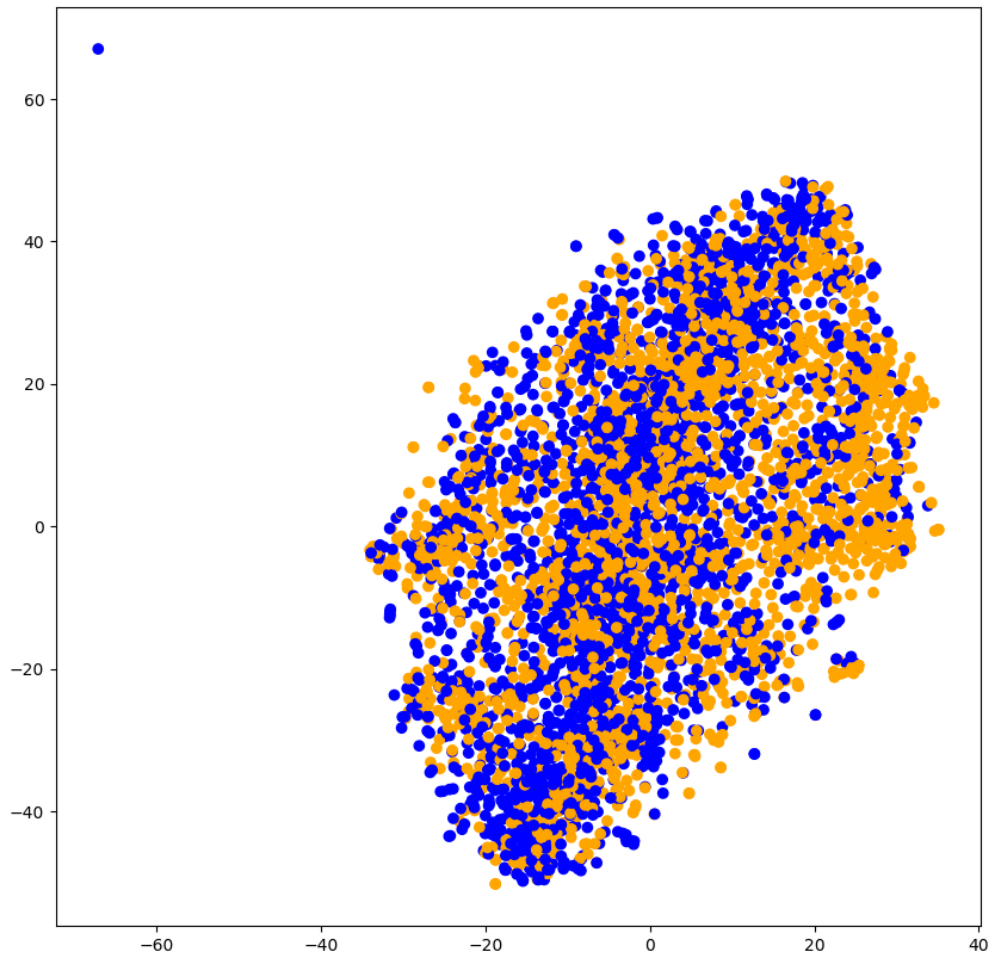Above **Logistic Regression**, below **Decision tree Classifier**.
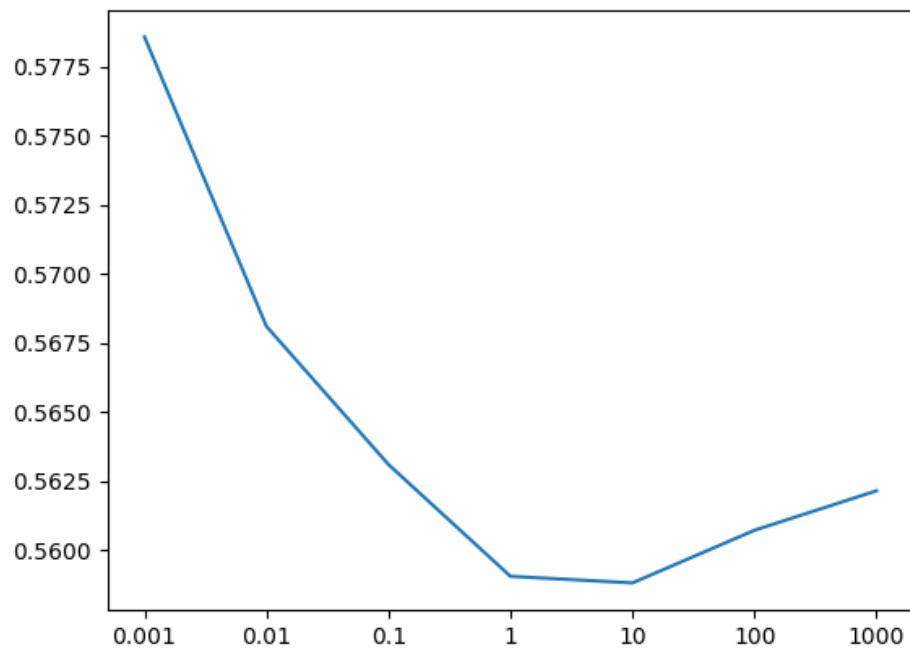

X-axis : Values of Parameter C
Y-axis: Values of Accuracy
As can be seen from the accuracy values and the plots above, **Logistic Regression** is the best model to predict class labels in this dataset.
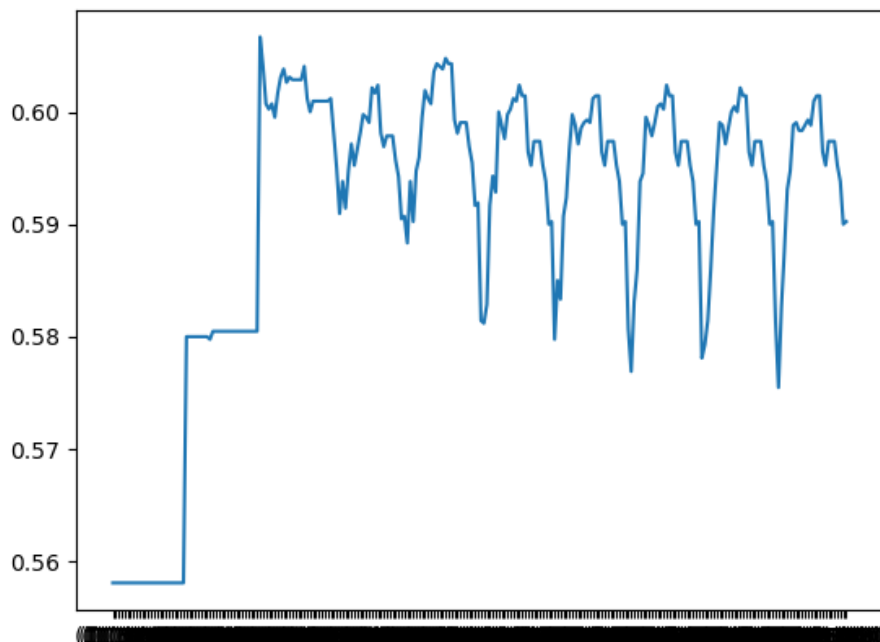
This is the visualisation for this dataset. This clearly has two classes.

Above **Logistic Regression**, below **Decision tree Classifier**.
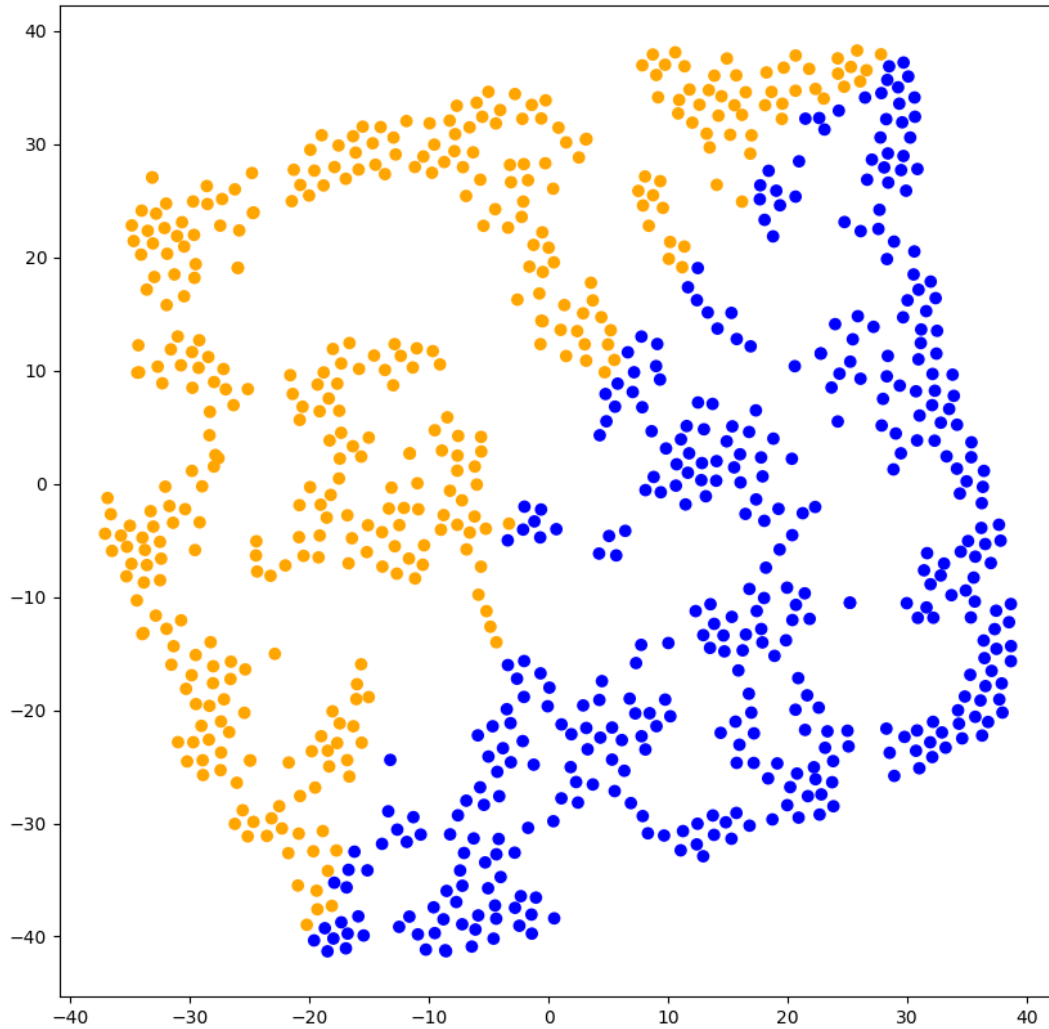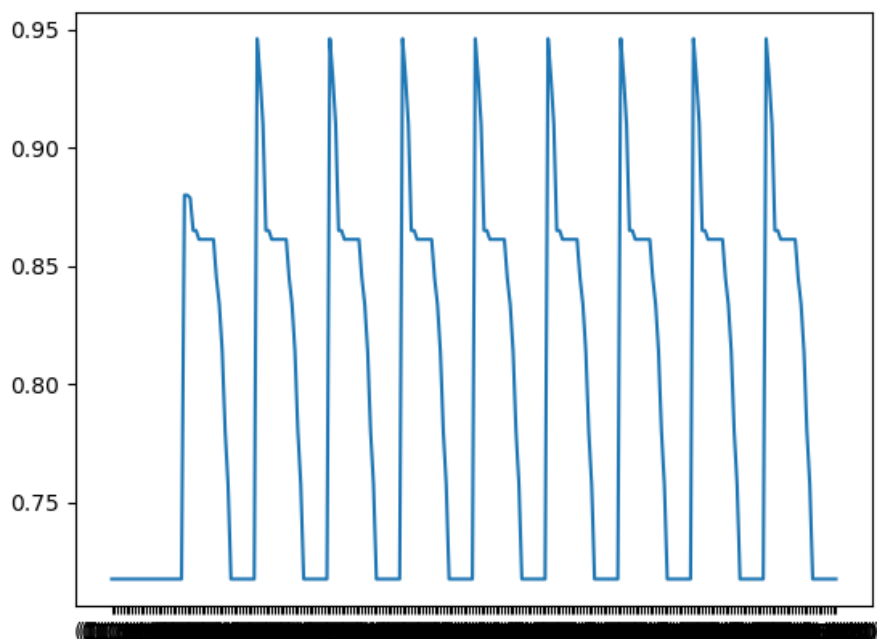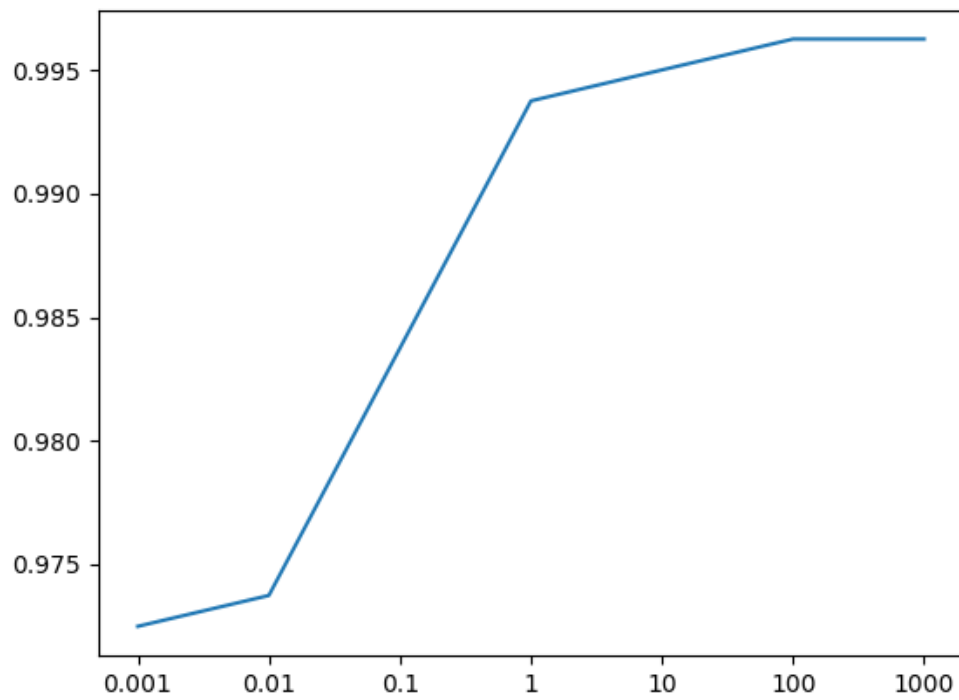

X-axis : Values of Hyperparameters
Y-axis: Values of Accuracy
As can be seen from the accuracy values and the plots above, **Decision Tree Classifier** is the best model to predict class labels in this dataset.

Dataset Part C
This is the visualisation for this dataset. This clearly has two classes.

Above **Logistic Regression**, below **Decision tree Classifier**.
-axis : Values of Hyperparameters
Y-axis: Values of Accuracy
As can be seen from the accuracy values and the plots above, **Logistic regression** is the best model to predict class labels in this dataset.