Muhammad Nadeem

201980050

DS-326-A-Machine Learning

Submitted to:

Sir Usman Ali

Assignment 1 Report

Problem Introduction:

In this Assignment We apply different Algorithms for Find the best accuracy on dataset. The data Those we give which has many files Firstly we Merge all files then we need to convert these files into 1400 rows and 1000 columns in which 1000 number column are called Labels. For convert Files into single file and Set rows and columns into given format.

Data Pre-Processing:

We Apply Pre-Processing on data to make a useful and understandable dataset. We select 14 files and collect 999 chunks from columns from 1 file and then select another 999 rows from another file. In which we find 1400 rows because we have 14 files and 999 columns because we collect 999 rows and convert into columns and 1000 column is a Label column. In which we assigned label to each row respectively. Finally, we have a file called Final.csv in which we have 1400 rows and 1000 columns. That's a Pre-Processing task.

PCA and KNN Implementation:

After pre processing we Apply PCA for find covariance matrix of data and then fin eigen vectors. Basically, PCA is used to reduce the dimensions of data. In Assignment we implement PCA Using Built in

Libraries and also explain From Scratch. After this we implement KNN using built in libraries and also implement from scratch. When we implement using built in libraries Model Accuracy is 11%. But when we implement from scratch our model accuracy is 25%. I which we also find Precision, Recall, F1 score, Confusion Matrix, Sensitivity and Specificity.

K-Fold Cross Validation:

Finally, We Apply K-Fold Cross Validation using KNN We find Average score of this model and plot graph for showing where we how many need folds for best accuracy.

In our bar chart graph, we need 6 folds for best Accurate prediction of model.