SEP 786 – Artificial Intelligence and Machine Learning Fundamentals

Project Results and Summary

Submitted by-

Neelanjan Goswami

400414867

Sumanth Mahabaleshwar Bhat

400351124

M.Eng Electrical and Computer Engineering

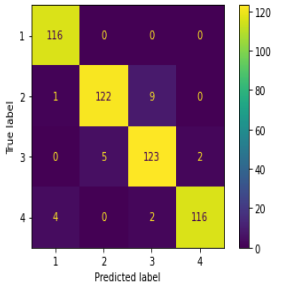
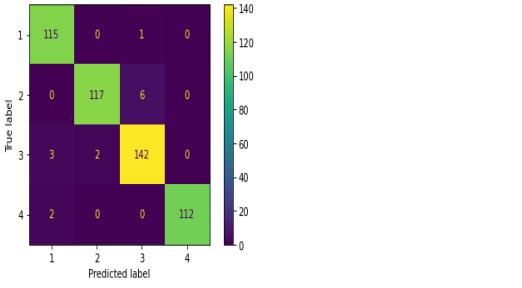
Faculty Advisor-

Jeff Fortuna

W Booth School of Engineering Practice and Technology

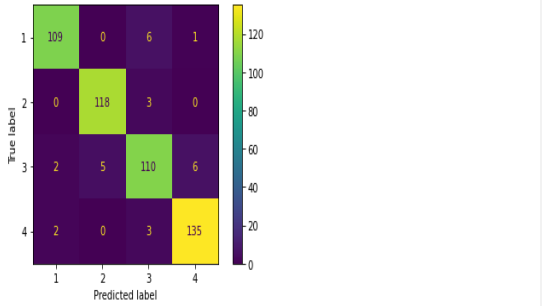
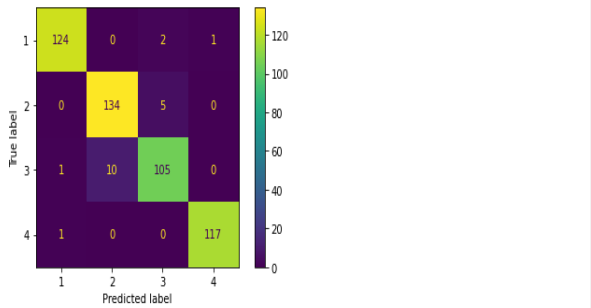
Results :

Logistic Regression

	PCA	Feature Selection
Accuracy	0.95	0.97
Number of Features	3	2
Training Time	0.033	0.287
Testing Time	0.003	0.007
Confusion Matrix	Confusion Matrix: 0 1 2 3 0 116 0 0 0 1 1 122 9 0 2 0 5 123 2 3 4 0 2 116 <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fc9c7fec6d0> 	Confusion Matrix: 0 1 2 3 0 115 0 1 0 1 0 117 6 0 2 3 2 142 0 3 2 0 0 112 <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fc9aee75d50> 

Decision Tree

	PCA	Feature Selection
Accuracy	0.94	0.96
Number of Features	3	2
Training Time	0.024	0.017
Testing Time	0.007	0.004

Confusion Matrix	Confusion Matrix: [[109 0 6 1] [0 118 3 0] [2 5 110 6] [2 0 3 135]] <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fc9af4f7d50> 	Confusion Matrix: [[124 0 2 1] [0 134 5 0] [1 10 105 0] [1 0 0 117]] <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fc9af36b150> 

Dataset :

(https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization#).

Wireless Indoor Localization Data, which is sourced from UCI ML, is the dataset utilised for this study. By analysing the WiFi signal levels in an interior setting, the data set is compiled. It is gathered by keeping an eye on the strengths shown on a smartphone. The dataset has one target variable and a total of seven characteristics. The WiFi signal strength as measured by a smartphone is each property. One of the four rooms is the goal value. Based on the seven signal intensities, the room is anticipated. There are a total of 2000 data points. 500 will be tested, and 1500 will be utilised for training.

Data Set Characteristics:	Multivariate	**Number of Instances:**	2000	**Area:**	Computer
Attribute Characteristics:	Real	**Number of Attributes:**	7	**Date Donated**	2017-12-04
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	67356

Summary :

On the data, we had to use PCA and Feature Selection. Logistic Regression and Decision Tree were the two classifiers we employed. The dataset was mean-centred and scaled for PCA. The Principle components were then determined using the Covariance matrix, Eigen Values, and Vectors. The scores for all of the principle components were determined using the Eigen Vector (principal components) and the data. Next, we developed models for all values of n (n = number of Principle components) to identify the number of Principle components to be employed.

We could detect which component the model was overfitting based on the accuracies. The model was overfitting after the third component in both Logistic Regression and Decision Tree. So we created both models for n = 3. The results were recorded (training time, testing time, accuracy, and confusion matrix). We used the backward search strategy to select the greatest to worst features for the Feature Selection. The data was organised from best to worst columns. Next, we developed models for all

values of n (n = number of features) to decide the number of features to be employed. We could assess how many characteristics the model was overfitting by looking at the accuracies.

The model was overfitting after two features in both Logistic Regression and Decision Tree. So we created both models for n = 2. The results were recorded (training time, testing time, accuracy, and confusion matrix). According to the data (shown in results.docx), Feature Selection outperforms PCA for both classifiers. The goal of PCA is to reduce dimensionality. We eliminate characteristics with a low variance. As a result, information will be lost. In addition, the PCA does not take into account the target variable in this technique. On the other hand, in Feature Selection, we remove the characteristic based on accuracy while keeping the desired value in mind.