ML PROJECT REPORT

Sachin Kumar – B19EE075
Shashi Prakash – B19EE076
Shivam Rathod – B19CSE083

INTRODUCTION

We trained 3 models ; every student trained one model . This document explains that three models used to classify images on Glass vs No Glass. All models are performed on the dataset provided. The models are performed on the data after dimension reduction also and in this way, there are 2 models per notebook that we have submitted, named respectively with roll numbers of students who trained that models.

DATASET

The topic of our project is **Glass vs no Glass classification**. The data set is composed of 4500 rows and 514 columns. Each column represents a picture . So, we can say that we have 4500 pictures that are converted to a data that python can interpret.

STEPS USED FOR ALL MODELS

1) *Data pre-processing* : First we preprocess the given data set and check the null values .Then we converted the categorical data into numerical data.

2)*Explanatory Analysis:* Mean , standard deviation , minimum , maximum and count of values for every column in dataset is described .

3) *Train test split:* Before training the model, we split the data set into train and test dataset to check the accuracy of model.

4) *Model selection:* Then we choose best model for our dataset.

5*) Train our model:* After that we fit our training data set in the model.

6) *Check the accuracy:* Then we checked the accuracy , cross validation score and computed confusion matrix and plotted ROC -AUC .

7) *Apply PCA* : Then we transformed our data into a smaller dimension data using Principle Component Analysis and do the same steps from 3 to 6.

A COUNT - PLOT OF CLASSES
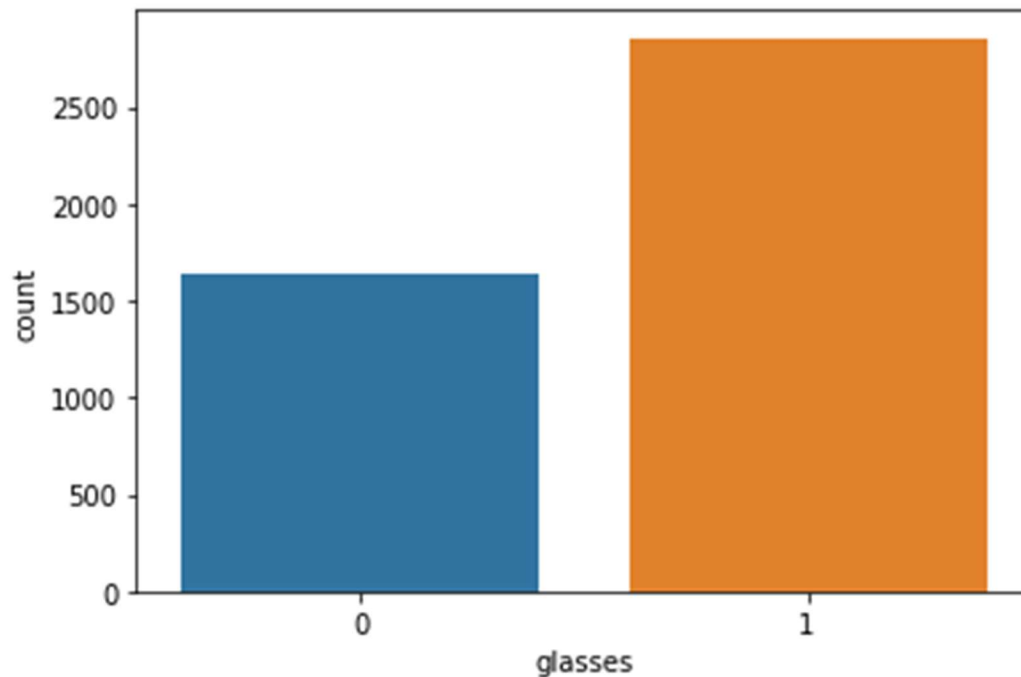
BRIEF DESCRIPTION OF ALL MODELS

MODEL 1 – LOGISTIC REGRESSION

This model is trained by the above steps mentioned by Sachin (B19EE075) and accuracy of this model is 98.89% and after dimension reduction by PCA its accuracy increased to 99.67%.

MODEL 2 – SVM MODEL

This model is trained by the above steps mentioned by Shashi Prakash (B19EE076) and accuracy of this model is 100% and after dimension reduction by PCA its accuracy remains same . So, it is clear that SVM is best model for this classification problem.
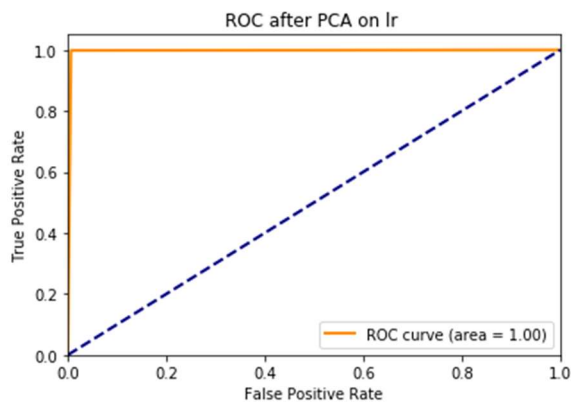
MODEL 3 – RANDOM FOREST CLASSIFIER

This model is trained by the above steps mentioned by Shivam Rathod (B19CSE083) and accuracy of this model is 90.55% and after dimension reduction by PCA its accuracy increased very significantly to an extent of 99.89% i.e. very good accuracy.

COMPARISION TABLE FOR ALL MODELS

MODEL	ACCURACY	CV – SCORE	CONFUSION MATRIX
LOGISTIC REGRESSION	98.89%	[0.9202 0.92 0.919]	[[298 10] [0 592]]
LOGISTIC REGRESSION AFTER PCA	99.67%	[0.970 0.9633 0.9799]	[[306 2 [1 591]]
SVM	100%	[1 1 1]	[[331 0] [0 569]]
SVM AFTER PCA	100%	[1 1 1]	[[331 0] [0 569]]
RANDOM FOREST	90.55%	[0.80 0.7866 0.765]	[[257 51] [34 558]]
RANDOM FOREST AFTER PCA	99.89%	[0.9966 0.99 0.9799]	[[308 0] [1 591]

ROC OF LOGISTIC REGRESSION
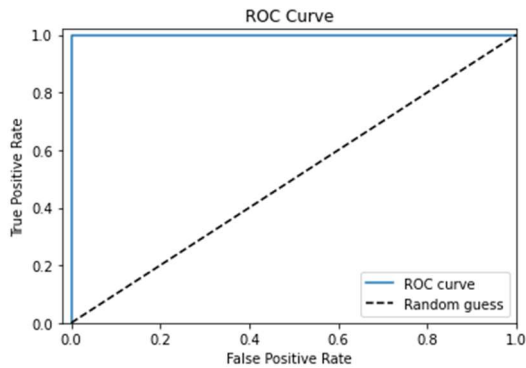

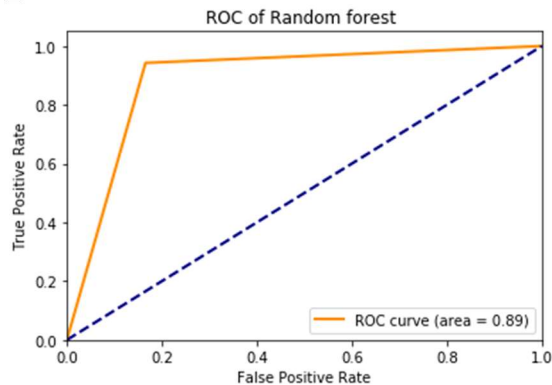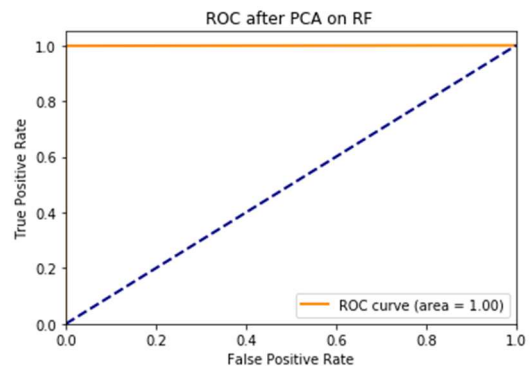With PCA
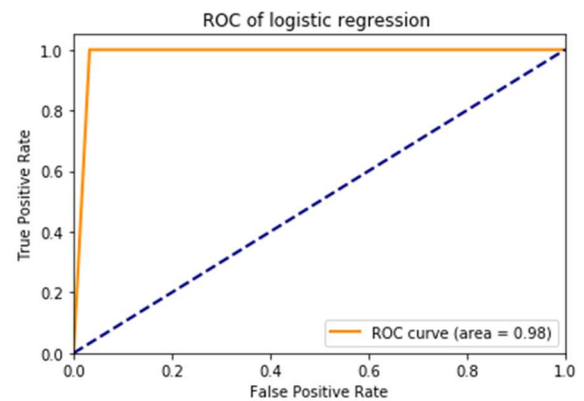
Without PCA

ROC OF SVM

❖ Both ROCs are same as its accuracy is 100% both models.

ROC OF RANDOM FOREST CLASSIFIER

=

Without PCA

With PCA

CONCLUSION

1) We classified our data using 6 different models having their own different accuracies and finally concluded that SVM model is the best model for classification of "**Glass Vs NO Glass**" data

.

2) SVM doesn't require any dimension reduction , its accuracy is 100% without reducing the size