Brain Computer Interaction

Final project

Eye state classification using Electroencephalography (EEG)

GROUP MEMBERS

D.vishal. S20190020210

Fawda Mizra S20190010050

1.) **INTRODUCTION**

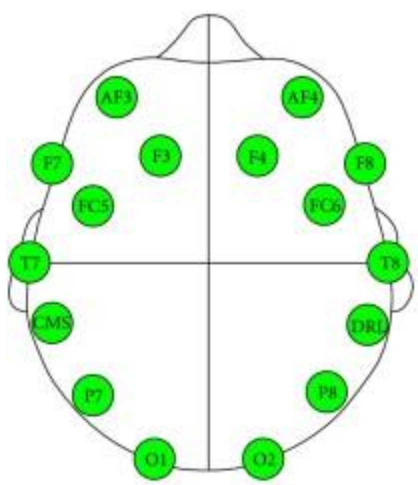
Electroencephalography (EEG) is used to measure the brain activity which can be used as inputs for applications requiring human actions. For example, brain stimuli have been used for computer games for handicapped persons etc. So, these require accurate detection of the stimuli to avoid false predictions. Many studies have been done to find differences between the two eye states, whether eyes are open or closed concluded that the greatest difference between two states was that the power in the eye closed state was much higher than that in the eye open state investigated how to track eye blinking based on EEG input.

2.) **Problem Statement**

The goal is to see if EEG signals obtained from the brain via various sensors can be used to predict eye state

3.) **Data Extraction**

The duration of the measurement was 117 seconds. Both open or partially open eyes were categorized as open, only completely closed eyes were categorized as closed. Sensors could be split into two groups. In the first group, the maximum increases when eyes open while, in the other group, the minimum decreases in the same event. Most sensors of the first group happen to be located on the right hemisphere while 1 most of the second group are on the left hemisphere of the brain, as displayed in Figure below



4.) **Data Cleaning**

i. We have assigned labels to the dataset, which helps to understand the dataset clearly
ii. We searched for null or duplicate values in the dataset, there are no such values in the dataset.
iii. There are total 14980 instances, 14 features and 1 column representing the class
iv. An outlier is an object that deviates significantly from the rest of the objects
v. After assigning labels and removing null or duplicate values we should remove the outliers.
vi. We deleted the outliers presented in the dataset using three processes (Z-Test, Using IQR, Outliers)

5.) **Z-Test**

i. For checking whether the data is having normal distribution or not, we performed Z-test using QQ-plot
ii. So we found z-score to each instance and the data having score less than 3 is considered as required data and the following analysis is applied. After removing the unwanted data there are 14892 instances

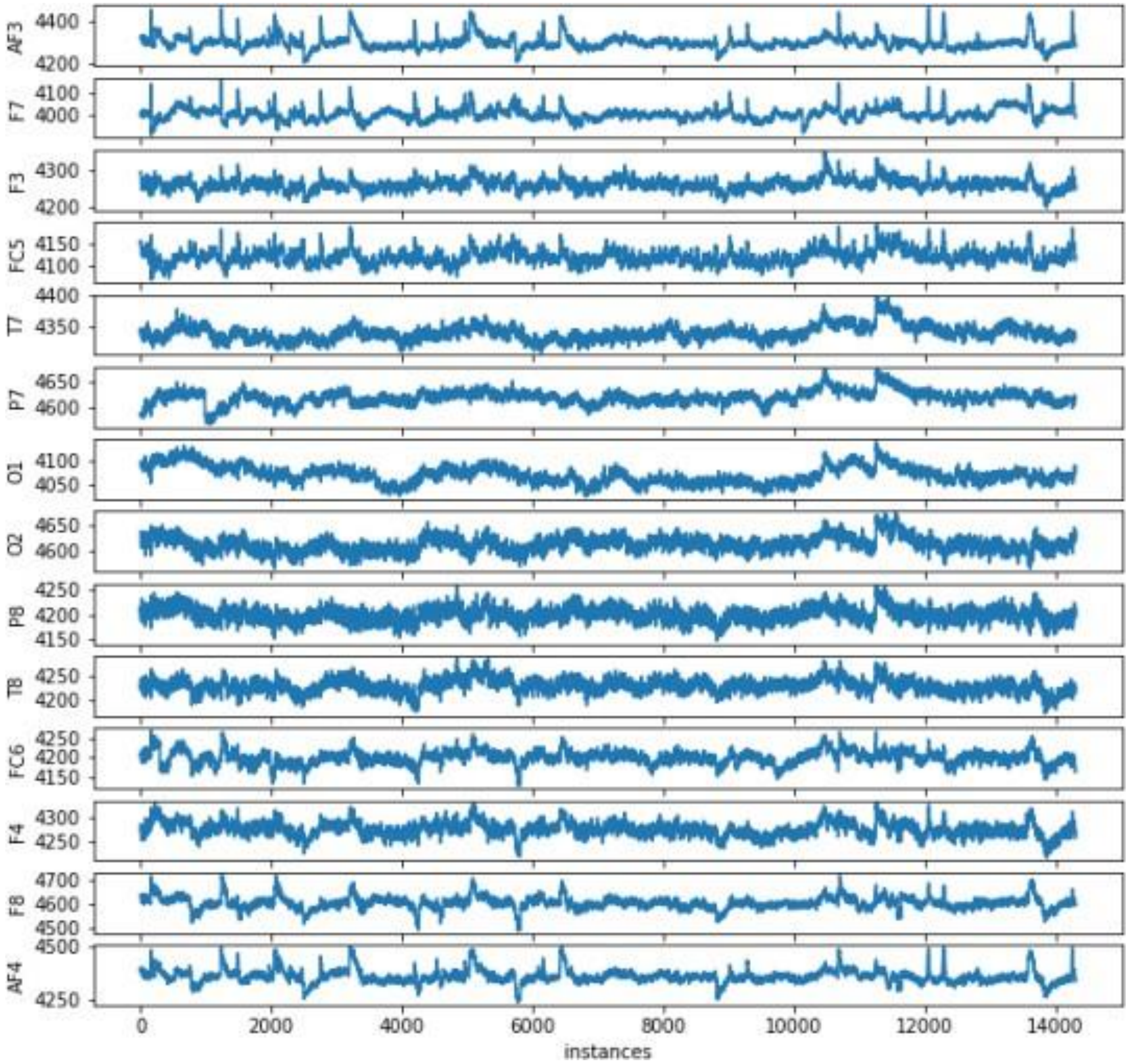
6.) **Using IQR**

i. In this we have found the quartiles (Q1,Q3) for each feature and also the inter quartile range(1QR = Q3-Q1). Now the data in between 2 the range ((Q1 - 3*IQR)),(Q3 + 3*IQR)) is considered as necessary data and following analysis is applied.
ii. After removing these instances there are 13865 data instances

7.) **Outlier**

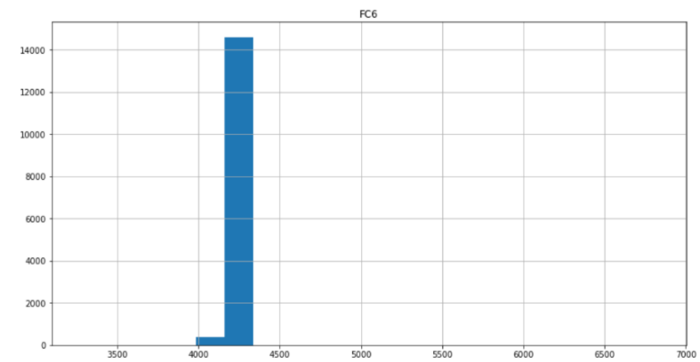
i. For this we have choose 4 times the standard deviation of the feature as the threshold. So whatever the data lies in (mean - std*4, mean + std*4) range, it will be considered as actual data.
ii. Now after removing the outliers the instances remained are 14304. So a total of 676 instances were removed

8.) **Visualizing sensor data:**

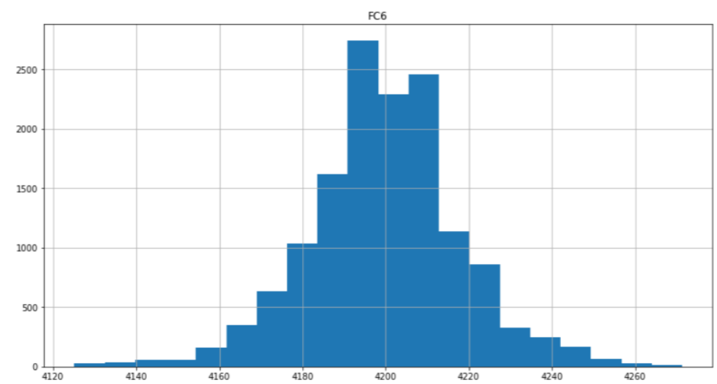


In the below figure we can see, even the distribution is now normal. This type of data will create dependent relations between the features.

1.) **Before removing outliers**



2.) **After removing outliers**

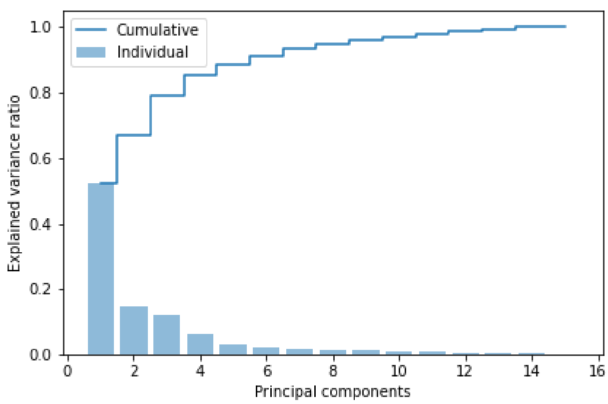


So after we applied statistical test for dimensionality reduction.

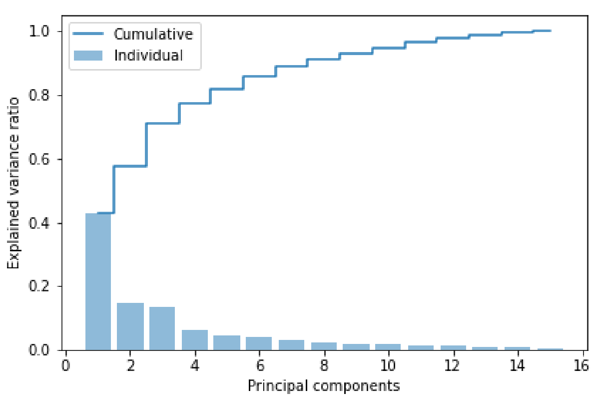
8) Principal Component Analysis (PCA)

i. Principal component analysis: PCA uses an orthogonal transformation to convert a set of observations of correlated variables into a set of linearly uncorrelated variables. These are called Principal Components.
ii. First covariance matrix is computed. Then eigen decomposition is performed on the covariance matrix and eigenvalues and vectors are computed. the PC’s were sorted out based on the corresponding eigenvalues. And then the cumulative sum is computed and plotted as shown below

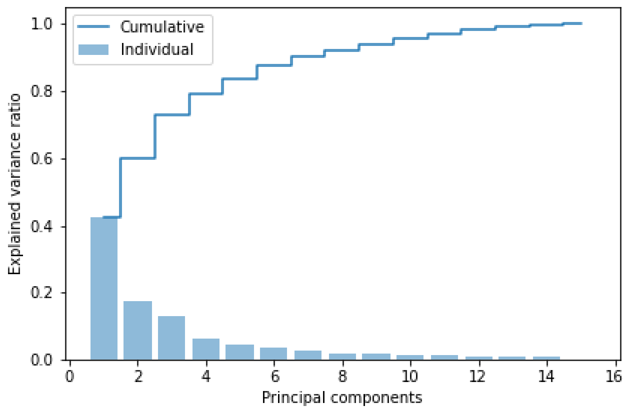
Filtered data using z-test:



Filtered data using IQR:



Without outliers:

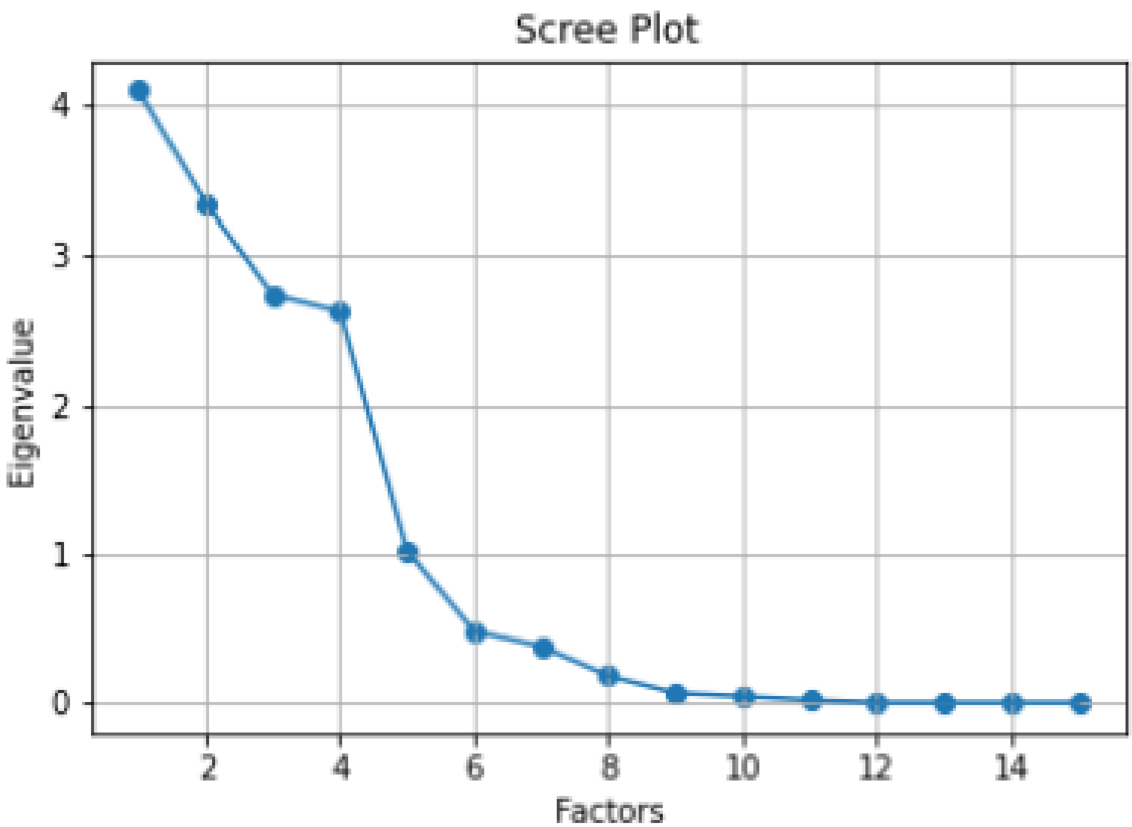


- Data filtered using z-test and IQR has 95% variance explained within the first 9 principal components and 11 principal components respectively.
- 95% of the variance is within the first 10 principal components for the data without outliers.

9) **Factor analysis:**

i. A factor is a hidden variable which describes the association among the number of observed variables. They explain the variance of each variable
ii. So we performed adequacy test using Kaiser-Meyer-Olkin Test, and the score is 0.856, which is excellent. It tells that there are hidden factors which describes the variables.
iii. Now factor analysis is applied for 6 factors.
iv. Factor 1 has high factor loadings for AF3, P8 and F8.
v. Factor 2 has high factor loadings for FC5 and O1.
vi. Factor 3 has high factor loadings for F3, P7 and AF4.
vii. Factor 4 has high factor loadings for F7, O2 and T8.
viii. Factor 5 has high factor loadings for T7 and FC6

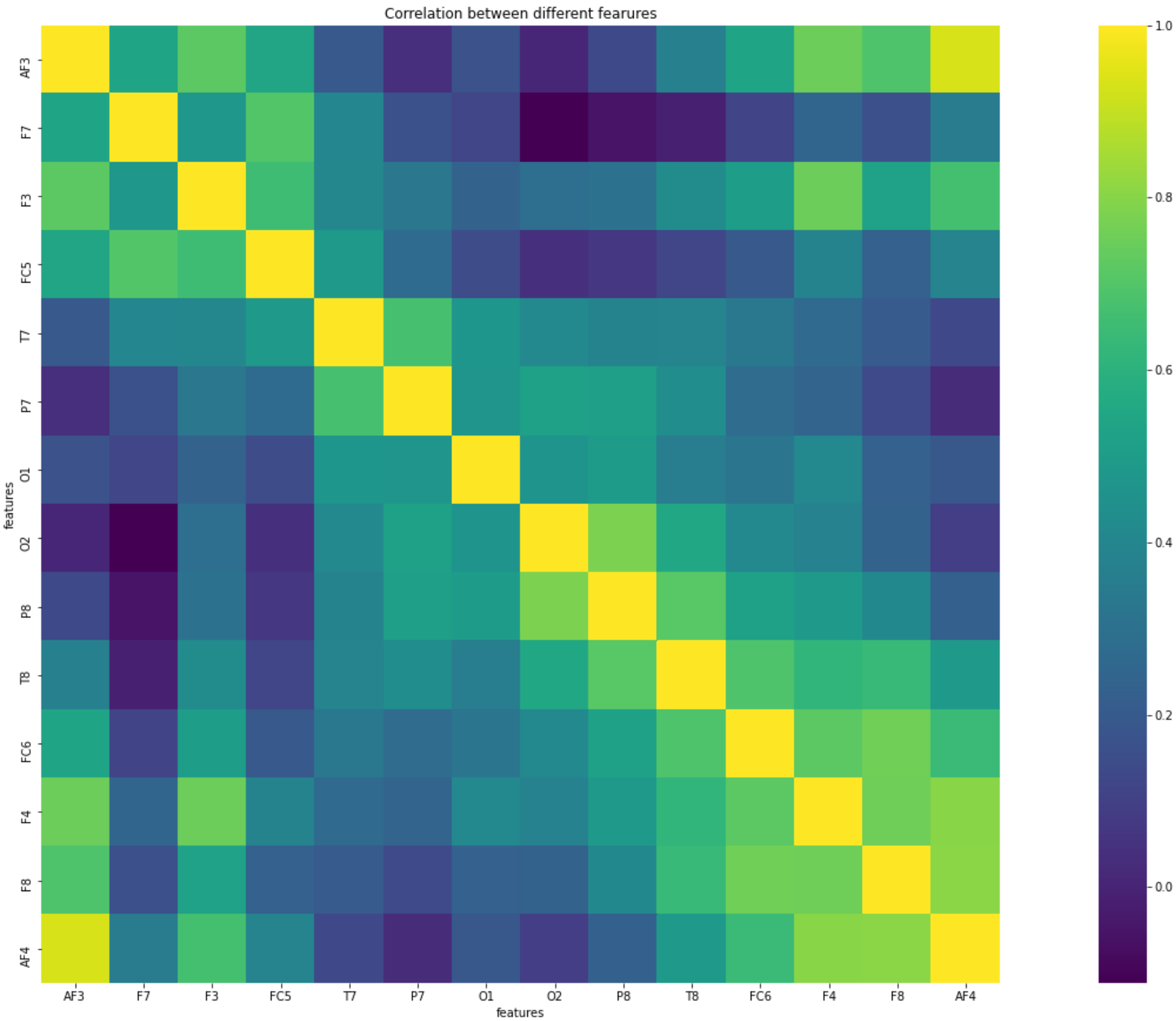
- **Screen plot for actual data:**



In the above figure we can observe an elbow at 5 number of factors. It represents that 5 factors are explaining the variance of all the features

10.) **Correlation heatmap:**

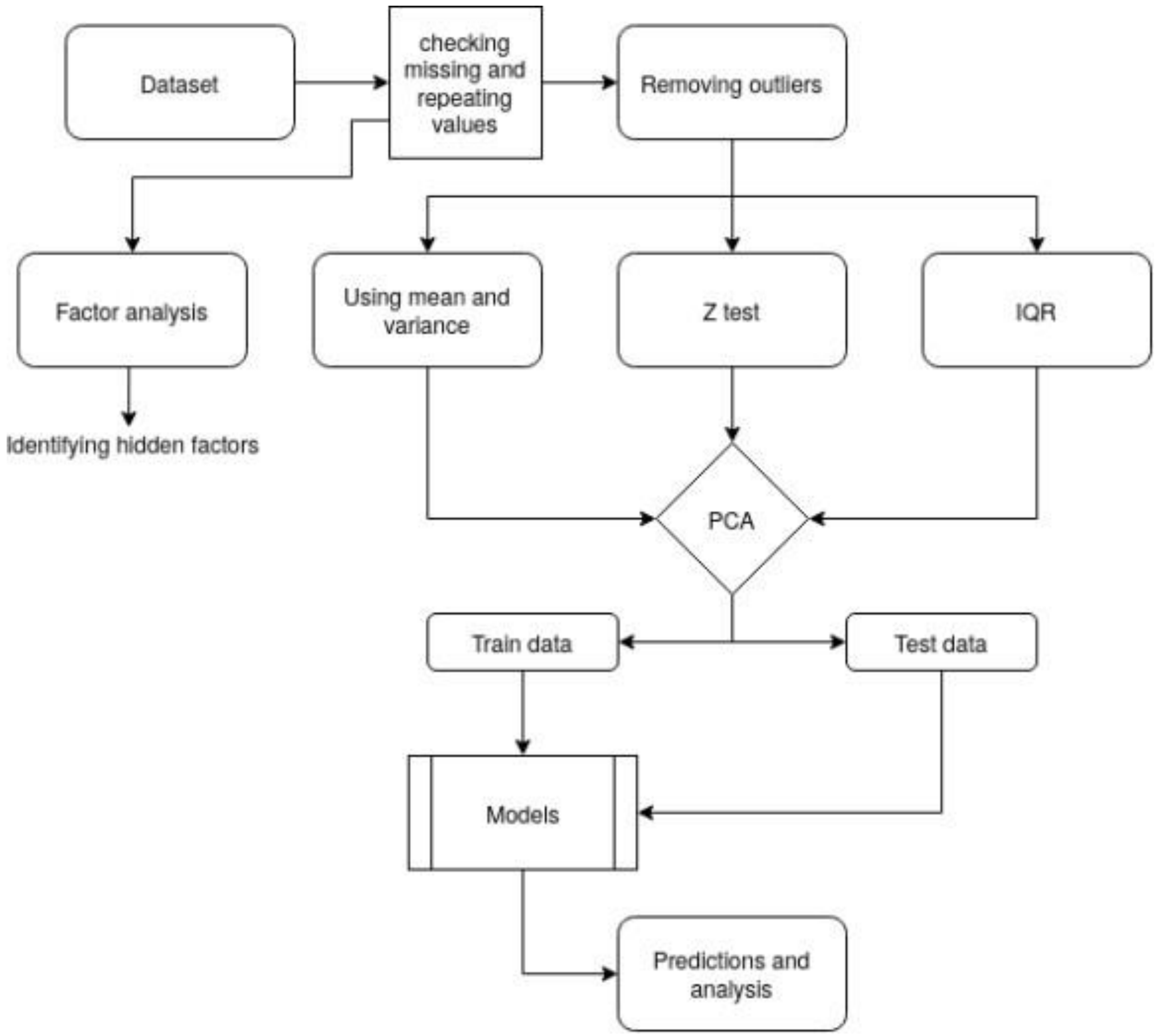
- For the actual dataset, correlation heatmap is plotted which can be seen in below figure.



The yellow color represents correlation of 1 and navy blue represents less than 0

- The diagonal matrix represents correlation between features vs the same feature. So it is having correlation of 1.
- Whereas most of the features has either very less or no correlation with other.
- Features such as T8, FC6, F4 and F8 have moderate correlation of around 0.4 to 0.7 with AF4.

11.) **Block Diagram:**

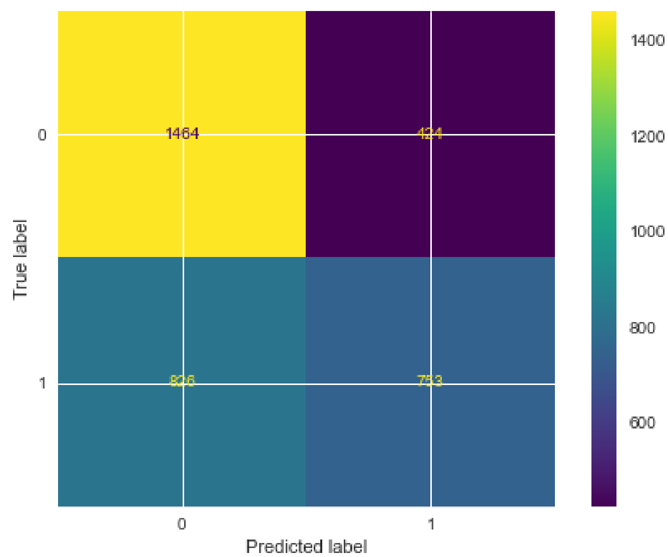


12.) Results:

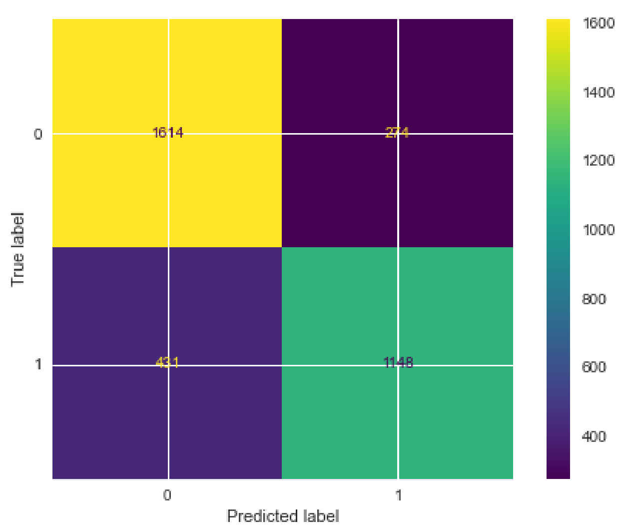
- For our classification experiments, we used sklearn library for implementing machine learning models. All the datasets, models such as LDA, QDA, SVM-gaussian, SVM-polynomial, Neural Networks are applied

- Linear Discriminant Analysis (LDA): LDA is a classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Bayes' rule

- Confusion matrix for LDA

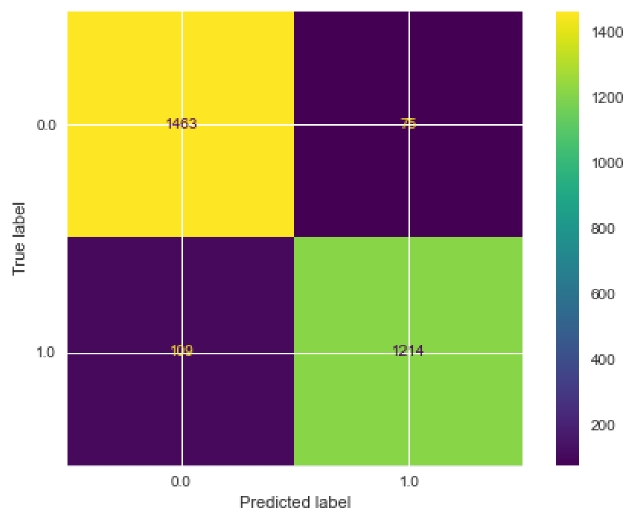


- Quadratic Discriminant Analysis (QDA): QDA is a variant of LDA in which an individual covariance matrix is estimated for every class of observations

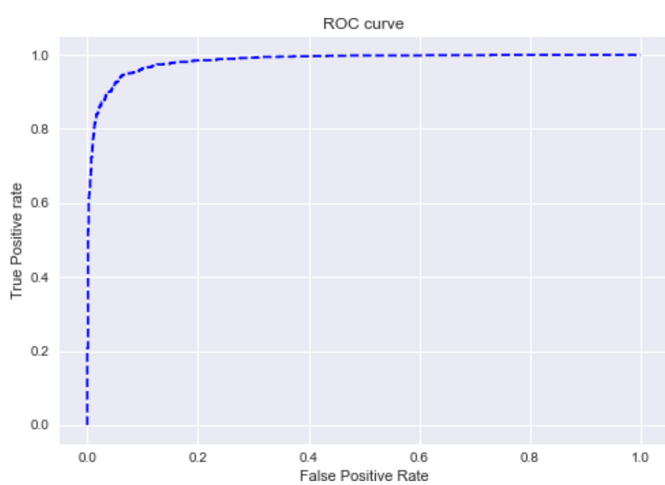


- Support Vector Machine (SVM): This algorithms use a set of mathematical functions that are defined as the kernel. The function of the kernel is to take data as input and transform it into the required form. SVM algorithm with different types of kernel functions have been implemented. Gaussian is really effective in higher dimension, polynomial kernel with degree 3 is implemented. The ROC value observed when SVM with gaussian kernel implemented is 98.3.

- **Confusion matrix for SVM:**



ROC:



13.) **Analysis:**

- accuracy for different models for the **raw data** as shown in thebelow table

S.No	Model	Accuracy
1	LDA	64%
2	QDA	50%
3	SVM-Poly	55%
4	SVM-Gaussian	55%

- Below table shows the performance of different estimators for all the datasets using different filtering processes, with and without PCA can be seen in below table. So the number of features used without PCA are 14.
- Whereas with PCA, the data without outliers has 10 features, data filtered using z-test has 11 features and data filtered using 1QR has 9 features.

- **With PCA:**

S.No	Model	Z-Test	IQR	Outliers
1	LDA	61%	65%	63%
2	QDA	73%	76%	74%
3	SVM-Poly	79%	87%	82%
4	SVM-Gaussian	92%	96%	94%

- **Without PCA:**

S.No	Model	Z-Test	IQR	Outliers
1	LDA	64%	64%	64%
2	QDA	78%	80%	79%
3	SVM-Poly	66%	65%	65%
4	SVM-Gaussian	62%	61%	60%

- From the above results of estimators, feature selection based on PCA has shown better accuracy results when compared to without dimensional reduction.

14.) **Conclusion:**

- We can conclude that it is possible to predict eye state using EEG sensor input with an accuracy of around 94%.
- Among all the classifiers SVM performed better after applying various filtering and dimension reduction techniques.

15.) **References:**

- P. Pour, T. Gulrez, O. AlZoubi, G. Gargiulo, and R. Calvo, “BrainComputer Interface: Next Generation Thought Controlled Distributed Video Game Development Platform,” in Proc. of the CIG, Perth, Australia, 2008.
- L. Li, L. Xiao, and L. Chen, “Differences of EEG between Eyes-Open and Eyes-Closed States Based on Autoregressive Method,” Journal Of Electronic Science And Technology Of China,, vol. 7, no. 2, 2009.
- B. Chambayil, R. Singla, and R. Jha, “EEG Eye Blink Classification Using Neural Network,” in Proc. of the World Congress on Engineering, London, UK, 2010

16.) **Individual contribution:**

Vishal. D –S20190020210

Cleaning data (*Z-test and IQR*), PCA analysis and classification algorithms . Results and comparison.

Fawad Mizra - S20190010050

Cleaning data (*outliers*), analysis before PCA and classification algorithms for outliers.

Analysis for raw data, theory and references