

## DATA MINING ASSIGNMENT #3

### NAMES:

- 1) ROHITH SARANGA (ID: 2017A7PS0034H)
- 2) KAUSHIK PERIKA (ID: 2017A7PS0207H)
- 3) MAHESH BABU (ID: 2017A7PS0235H)
- 4) BHARGAV NIMMAGADDA (ID: 2017A7PS1574H)

**DATASET USED:** CREDIT CARD FRAUD DETECTION (link: <https://www.kaggle.com/mlg-ulb/creditcardfraud>)

### PRE-PROCESSING:

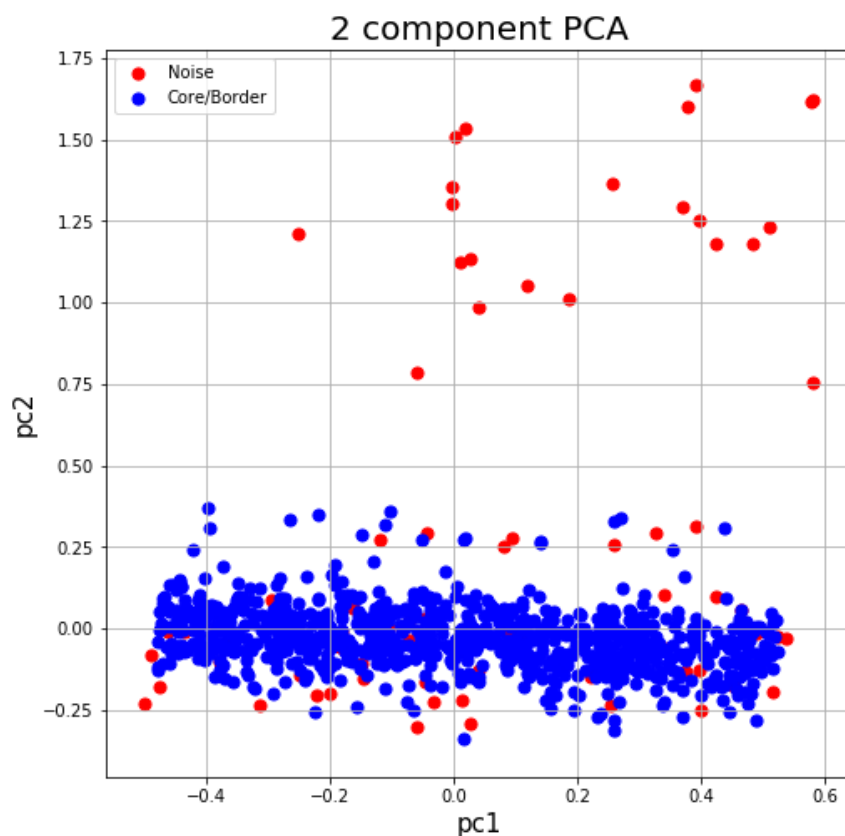
The dataset consisted of 2,85,000 rows and 31 columns. Since working with such large data would be computationally intensive we decided to take only a sample of 1000 rows which had 22 rows with classes 1 (denoting noise points) and others of class 0 (denoting normal points).

### PLOTS:

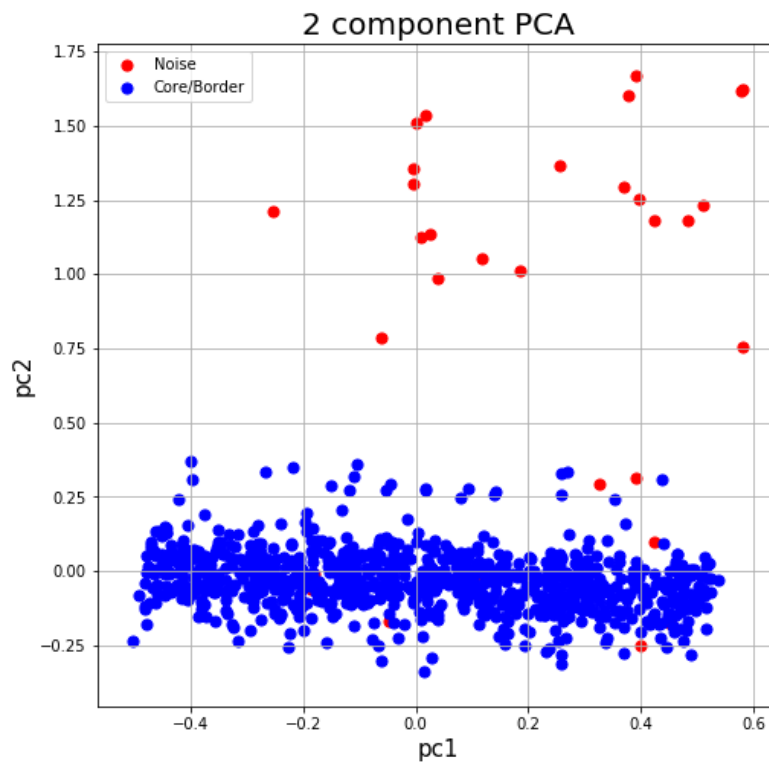
PCA (Principal Component Analysis) has been applied to the data set to transform the data into 2D and we made use of sklearn. PCA (scikit learn library) class to plot the image of outliers.

1)DBSCAN:-

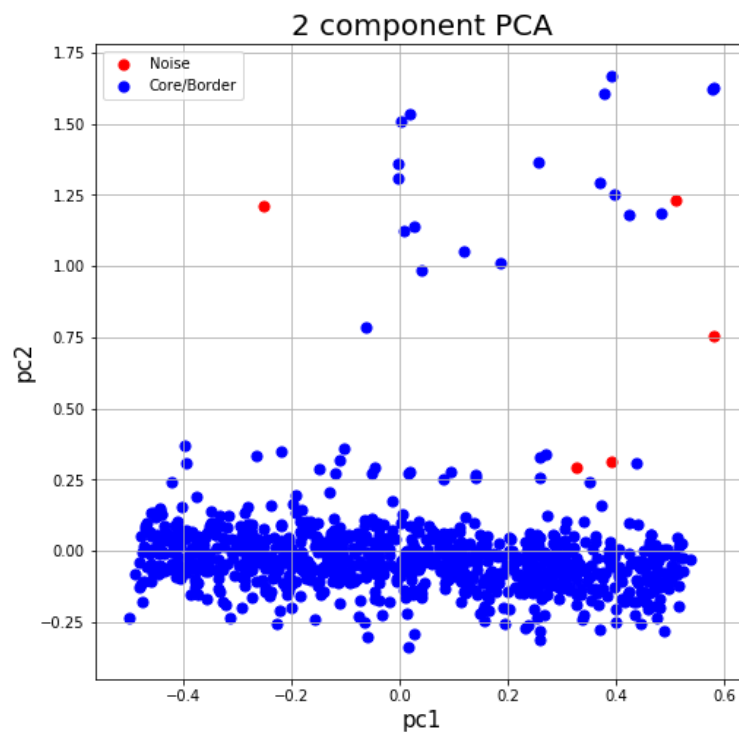
eps = 2 Accuracy = 94.0



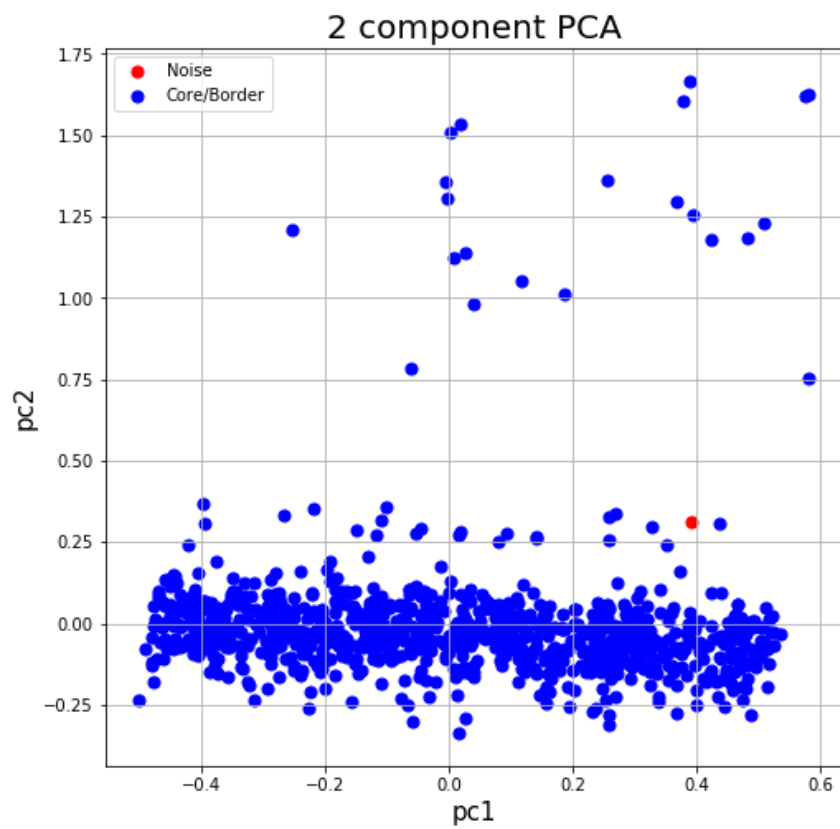
eps = 3 Accuracy = 98.8



eps = 4 Accuracy = 97.8

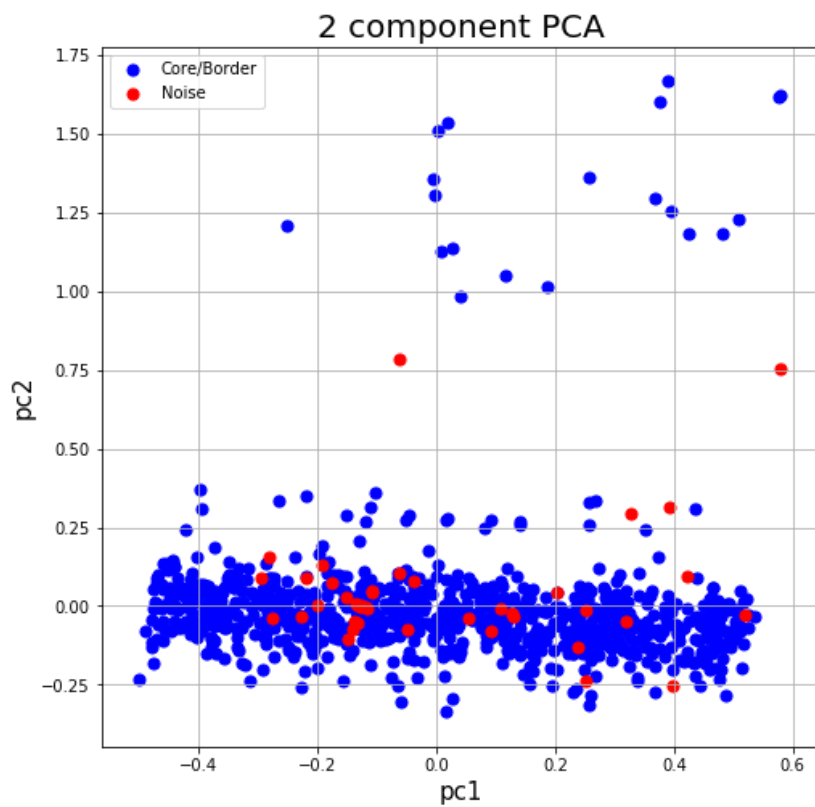


eps = 8 Accuracy = 97.7

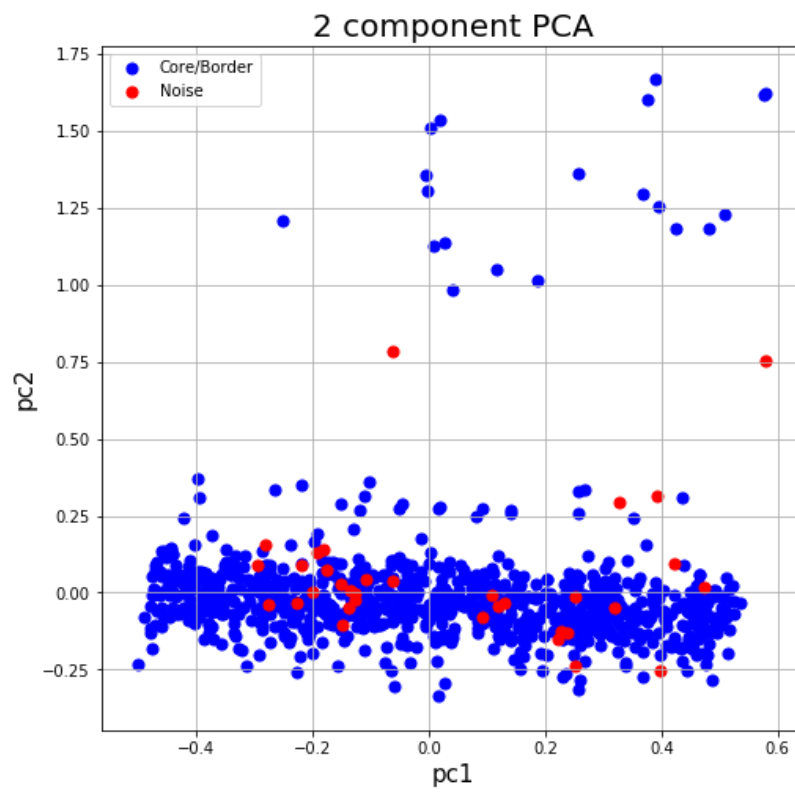


## LOF

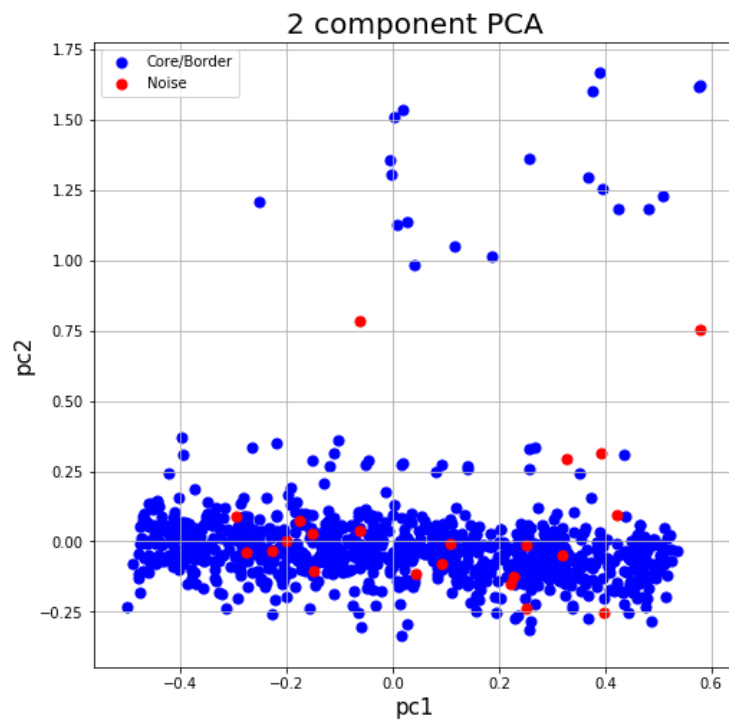
k = 4 Accuracy = 94.1



k = 5 Accuracy = 94.8



K = 6 Accuracy = 96.0



k = 7 Accuracy = 96.0

