CS6890 - Fraud Analytics - Assignment 3

Tanmay Garg
CS20BTECH11063

Aayush Patel
CS20BTECH11001

Shashank Shanbhag
CS20BTECH11061

Ganesh Bombatkar
cs20btech11016

Sushma
cs20btech11051

Abstract—**This document is a report of the implementation of outlier detection using Spectral Clustering on dealer transaction data.**

The libraries used are `Scipy`, `Sklearn`, `NumPy`, `Pandas` and `Matplotlib`.

Index Terms—**Outlier Detection, Spectral Clustering, Fraud Analytics, Eigenvalues, EigenVectors, K-Means Clustering, Laplacian Matrix**

I. INTRODUCTION

II. PROBLEM STATEMENT

III. DATA DESCRIPTION

In given dataset, there are 1999 rows and 10 columns. Each row represents a transaction made by a dealer. Columns are feature derived from the raw data. The features are as follows:

	mean	std	min	max
cov1	0.96	0.14	-0.31	1.00
cov2	0.86	0.24	-0.53	1.00
cov3	0.21	0.41	-0.82	1.00
cov4	0.15	0.39	-0.84	0.98
cov5	0.04	0.18	-0.72	1.00
cov6	0.60	0.33	-0.68	1.00
cov7	0.53	0.39	-0.86	1.00
sal_pur_rat	-0.00	1.00	-0.04	34.37
igst_itc_tot_itc_rat	-0.00	1.00	-1.07	2.18
lib_igst_itc_rat	0.00	1.00	-0.05	33.19

Also the distribution of the data is shown in Figure **??**. which is derived from analysing the principal components of the data. It shows that data has one big cluster and some outliers.

IV. ALGORITHM

V. RESULTS

VI. COMPARISON

VII. CONCLUSION

Fig. 1: 2D representation of the data