ASSIGNMENT-REPORT

Data-Preprocessing

- Data-preprocessing is necessary because the dataset contains non-numeric values in some of the columns.It has been replaced by the mean of all the values of that particular column.

Data-Normalization

- Data normalization is necessary because the range of values in features 'Time' and 'Amount' is larger compared to other features.Min-Max scaler was used to normalize all the features.

RESULTS AND OBSERVATIONS

The clustering algorithms used were Kmeans and Birch .Kmeans algorithm was run on both normalized and un-normalized data.The following results were obtained and with Time feature and without Time feature.Birch Algorithm was run on normalized data with and without Time feature.

Algorithm	Accuracy Score	Matthew's Correlation Coefficient	Root Mean Square Error	Negative Predictive Value
Kmeans with normalization and 'Time' feature	0.53634	-0.0118	0.6809	0.3211
Kmeans without normalization and 'Time' feature	0.5369	-0.0111	0.6804	0.3292
Kmeans with normalization and 'Time' feature dropped	0.6775	0.0105689	0.56789	0.4410
Kmeans without normalization and 'Time' feature dropped	0.9801	0.0050	0.1409	0.03455
Birch with normalization and 'Time ' feature	0.9982	0.0	0.04155	0.0
Birch with normalization and 'Time ' feature dropped	0.9982	0.0	0.0415	0.0

INFERENCE AND CONCLUSION

Even though Birch has a higher accuracy score than Kmeans,Kmeans is actually much closer to actual result beacouse of its ability to cluster the minority class far more efficiently than Birch which is evident from its correlation coefficient.Kmeans with normalization and Time feature dropped is closer to the actual result as compared Kmeans without normalization and 'Time' feature dropped because of the former's higher negative predictive value than the latter's.