Data info is available in “Qualitative\_Bankruptcy.info.txt” file

Please follow the below steps to complete the project

1. Read the dataset and understand the features and target variable.
2. Encode independent features.
3. Separate the dependant variable from the rest of the train data.
4. Split the dataset into training and test data ( Take 80:20 ratio).
5. Apply PCA on the independent features of train data.
6. Display elbow plot to find out reduced number of dimensions (which covers at least 95% of the variance).
7. Transform the data using PCA ( select optimum number of PCs).
8. Apply KMeans algorithm on the above transformed data.
9. Select optimal number of clusters and form clusters out of the train data.
10. Create a new column in the train data and populate it with the cluster labels.
11. Transform the test data using the same PCA object that was fit on the train data.
12. Use the same kmeans object, assign cluster labels to the test data and store it as a feature.
13. Apply any classification ( i.e. SVC, Logistic) model on train dataset and predict target for test dataset.
14. Check accuracy and confusion matrix.