FINAL PROJECT – MACHINE LEARNING

Github link :

https://github.com/Nalinalingasamy/Machine-Learning-Algorithms

PROBLEM 1:

With the table containing the information of Customer price index, discounts, offers where the organization wants to predict the sales based on the cpi, discounts,offers.

For below information find out the Sales that has

5000 cpi , 3 percentage discounts, 20 rewards offers

4000 cpi , 8 percentage discounts, 19 rewards offers

SOLUTION:

By using LinearRegression(Linear Model), I have predicted the sales as following

- 5000 cpi , 3 percentage discounts, 20 rewards offers

  Predicted value is Rs. **826645.348**

- 4000 cpi , 8 percentage discounts, 19 rewards offers

  Predicted value is Rs. **732680.364**

PROBLEM 2:

With the table containing the bank details of customer – need to predict the how we can offer loan to customer based on the cybil score, Age, insurance, debit card , cards.

Steps followed:

- There is no multi-collinearity between the independent variables.
- After splitting the train and test datas, linear model is used to predict the cases.
- The model has been fitted by x-train and y-train datas.
- Then the prediction has been done using x-test data.
- Accuracy score has been checked for x and y test data.

SOLUTION:

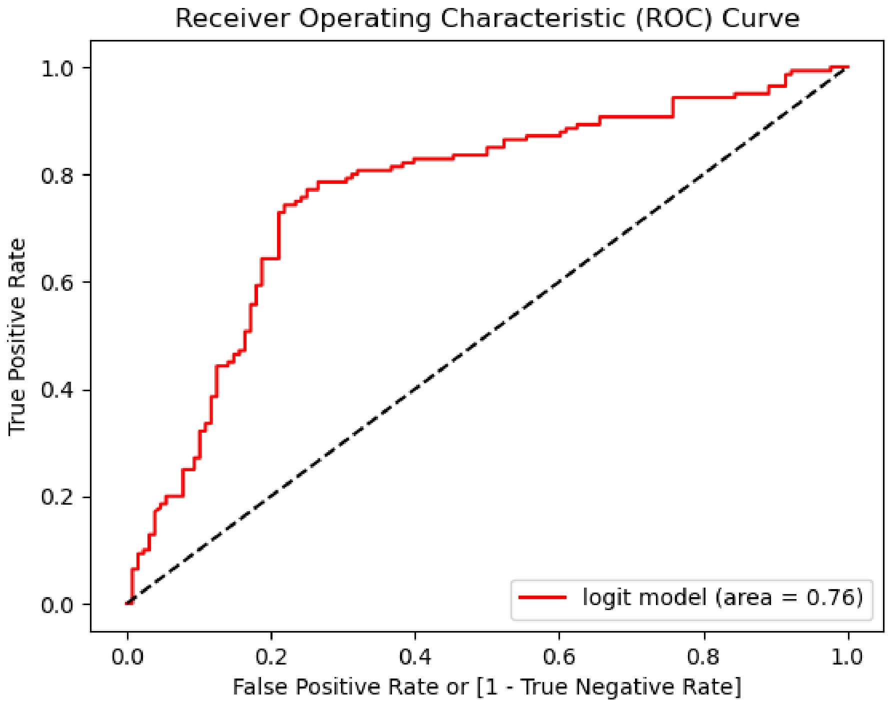
Following all the steps for building a model,

CLASSIFICATION REPORT IS AS FOLLOWS:

	PRECISION	**RECALL**	**F1 SCORE**
0	0.76	0.67	0.71
1	0.73	0.81	0.77
ACCURACY	0.74		

- Accuracy = **0.7425**
- Specificity (OR) True Negative Rate = **0.8071**
- Sensitivity (OR) recall (OR) true Positive Rate **= 0.6718**
- Precision = **0.7610**
- Area under curve(AUC) = **0.7647**

ROC CURVE :



PROBLEM 3:

With the table containing the information of customer and their geographical and personal details , Build the classification models like Decision Tree , Random forest , KNN , kmeans and SVM to classify the data in to proper groups.

SOLUTION:

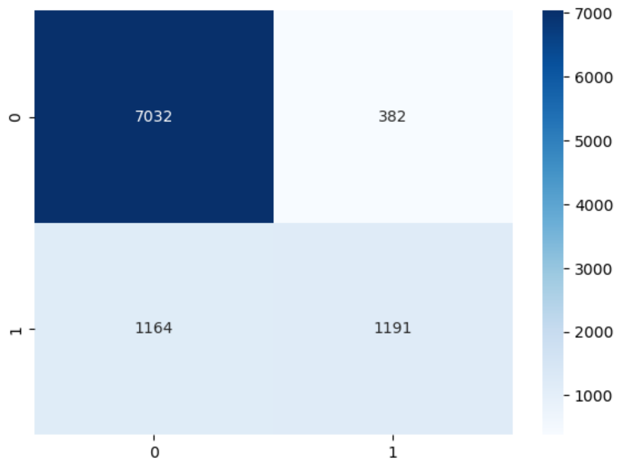
DATA PREPROCESSING STEPS:

- Null values have been found and treated (Null value imputation)
- Encoding has been done for categorical columns
- Data were splitted into train and test data

1.DECISION TREE CLASSIFIER:

- Model defined and fitted with train datas.
- Prediction was done with test datas.
- The classification report is as follows:

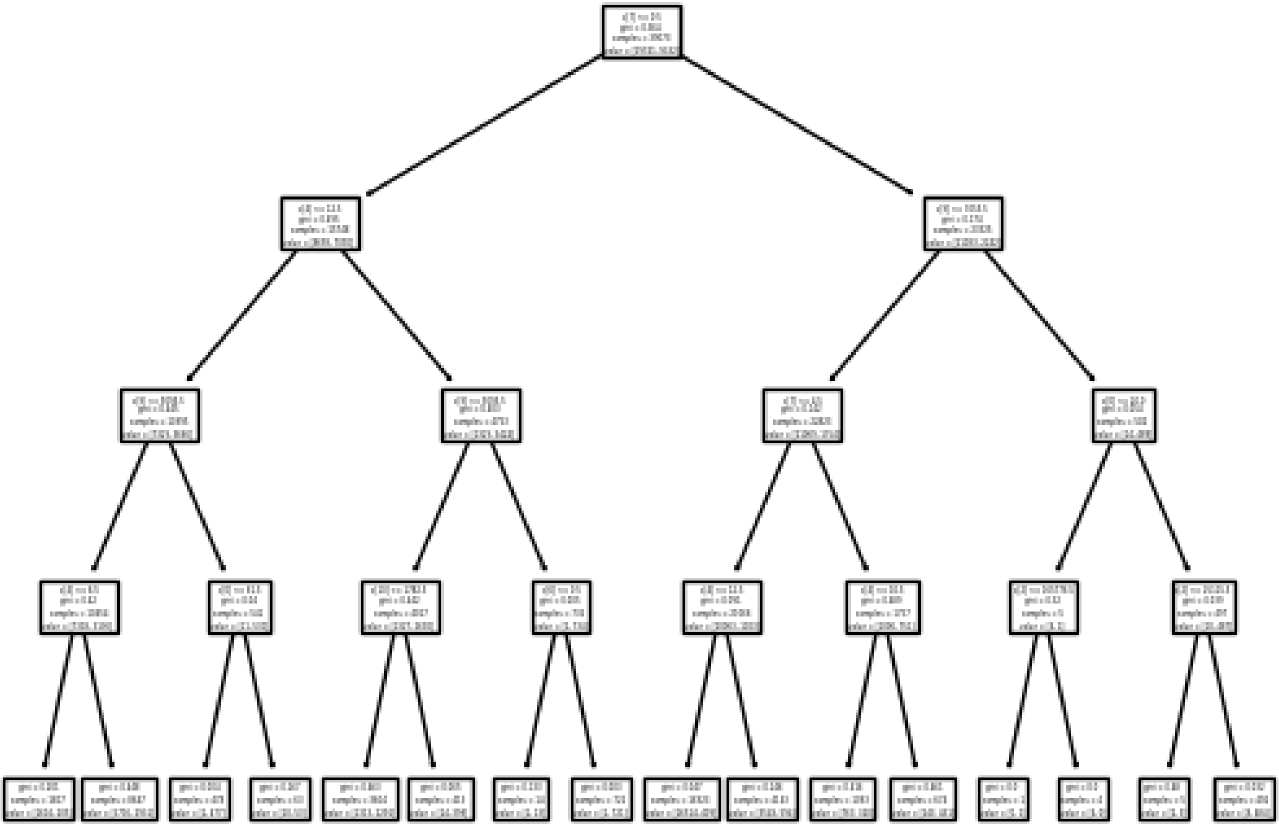
CONFUSION MATRIX



CLASSIFICATION REPORT:

	PRECISION	RECALL	F1 SCORE
0	0.86	0.95	0.90
1	0.76	0.51	0.61
ACCURACY	0.84		

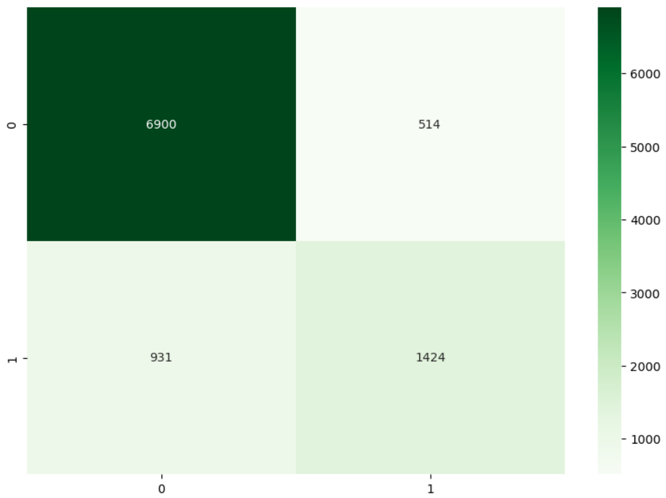
TREE FORMATION REPRESENTATION:



2.RANDOM FOREST CLASSIFIER:

- The classification report is as follows:

  CONFUSION MATRIX



-

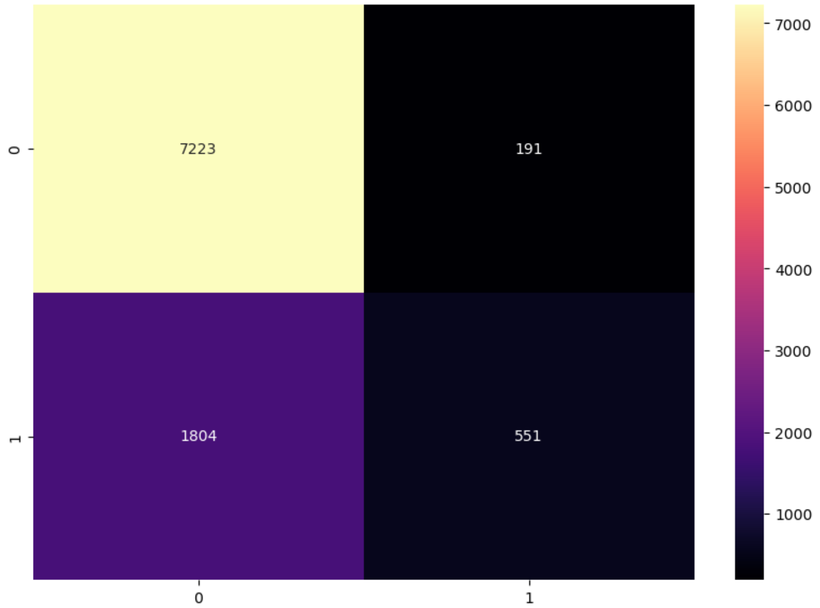
CLASSIFICATION REPORT:

	PRECISION	RECALL	F1 SCORE
0	0.88	0.93	0.91
1	0.73	0.60	0.66
ACCURACY	0.85		

3.K NEAREST NEIGHBOUR(KNN):

- The classification report is as follows:

CONFUSION MATRIX



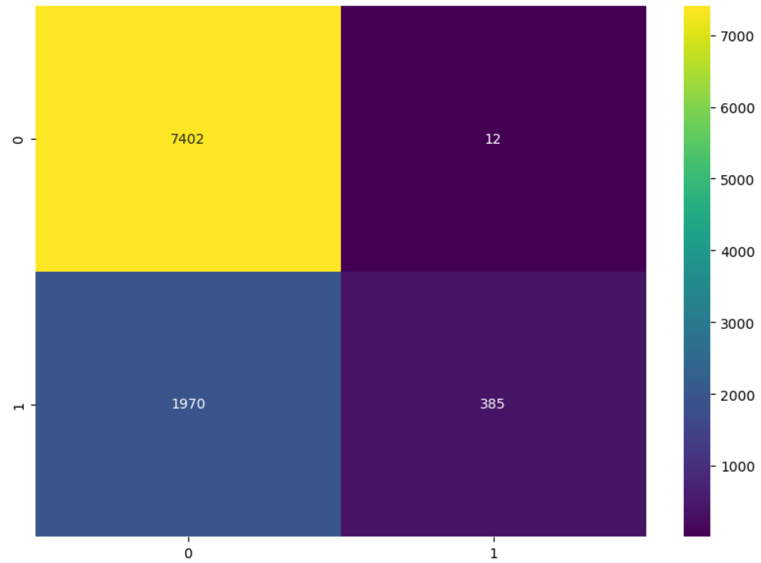
CLASSIFICATION REPORT:

	PRECISION	RECALL	F1 SCORE
0	0.80	0.97	0.88
1	0.74	0.23	0.36
ACCURACY	0.80		

4. SUPPORT VECTOR MACHINE:

- The classification report is as follows:

CONFUSION MATRIX



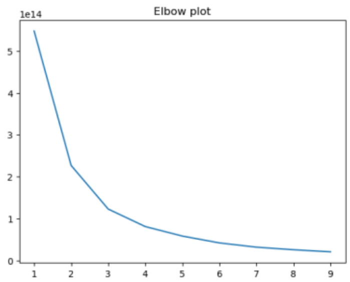
CLASSIFICATION REPORT:

	PRECISION	RECALL	F1 SCORE
0	0.79	1.00	0.88
1	0.97	0.16	0.28
ACCURACY	0.79		

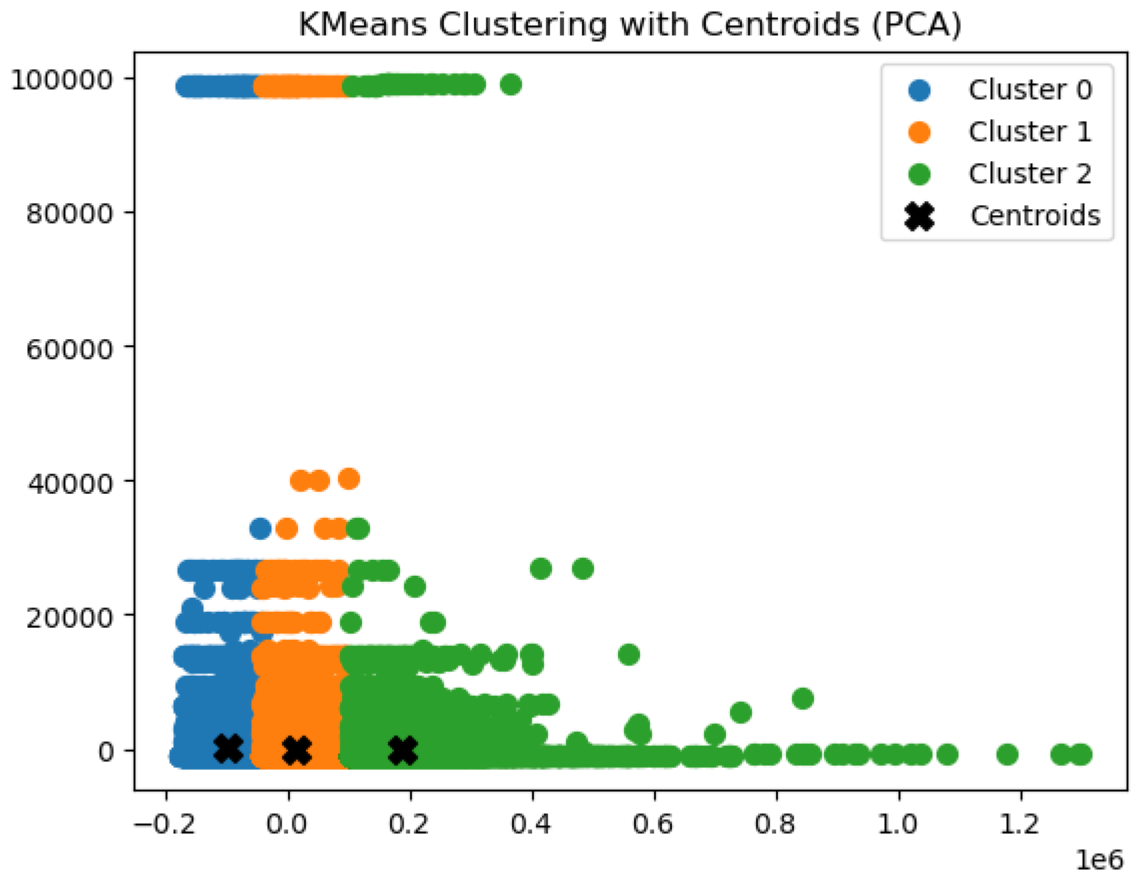
5.K-MEANS CLUSTERING:

- The optimal number of clusters was found by plotting elbow plot.
- Number of clusters determined to be **3**.

Elbow plot:



KMeans clustering graph with centroids:



By using "**silhouette score"** from **sklearn. metrics,** the effectiveness of the clustering,

Silhouette Score: 0.551

By evaluating all the models, the best model for this dataset was found to be **Random Forest Classifier** with an accuracy score of 85%.