

GLOSARRY

a, b – used in loop to find unique row identifier

amnt0 – mean value of balance where exited column values equal to 0

amnt1 – mean value of balance where exited column values equal to 1

acc - accuracy score

class1per - percentage of class 1 in Exited column

catcols09 - data type categorical columns

corr11 – correlation of numerical features

countbal – number of customers having less than average balance exited the bank

cleaned_dfm – main data frame after cleaned

conf_mtrx – confusion matrix

clf_report – classification report

confu_matrix – confusion matrix

cm_display – confusion matrix display

df_encoded – data frame after converting categorical features into dummies

dfm – main data frame

dupl – data frame with duplicate values

prod0 – count of unique products registered where exited column values equal to 0

prod1- count of unique products registered where exited column values equal to 1

rocauc – roc auc score

report – classification report

unq_valper – unique values percentage

Fig 1.1 First five rows of the data frame

Fig 1.2 Info of the data frame

Fig .13 Code snippet

Fig 1.4 Statistical summary of dataset

Fig 1.5 Summary of type categorical variables from dataset

Fig 1.6 Histogram plots of numerical variables

Fig 1.7 Distribution plot of Target variable

Fig 1.8 Histogram plots of Categorical variables

Fig 1.9 Code Snippet of unique values percentage of categorical variables

Fig 1.10 Code Snippet of unique values percentage of numerical variables

Fig 1.11 Pair plot of all the features

Fig 1.12 Spearman correlation table

Fig 1.13 Heatmap of Spearman correlation

Fig 1.14 Box plots of Age vs categorical variables

Fig 1.15 Scatter plot of Number of products registered vs Balance amount

Fig 2.1 Code snippet of encoded Data Frame

Fig 3.1 Snapshot of Libraries used for modelling

Fig 3.2 Snapshot of Sigmoid Function

Fig 3.3 Code snippet of Model build

Fig 3.4 Output of Predictive model

Fig 4.1 Code snippet of K fold cross validation and its output

Fig 4.2 Code snippet of SMOTE sampling technique and its output

Fig 4.3 Snapshot of Model output after SMOTE sampling

Fig 4.4 Heatmap of Confusion matrix

Fig 4.5 Snapshot of Feature importances for Model Random Forest classifier