BSC & MSC Data Science
Case Study :– Elective (Financial Analytics)
Marks :- 60
Date of Submission:- 11/03/2023

Website:
sdbi.in

Contact:
9920053700

Email:
info@sdbi.in

CONTEXT: A bank X is on a massive digital transformation for all its departments. Bank has a growing customer base majority of them are liability customers (depositors) vs borrowers (asset customers). The bank is interested in expanding the borrowers base rapidly to bring in more business via loan interests. A campaign that the bank ran in last quarter showed an average single digit conversion rate. Digital transformation being the core strength of the business strategy, marketing department wants to devise effective campaigns with better target marketing to increase the conversion ratio to double digit with same budget as per last campaign.

DATA DESCRIPTION: The data consists of the following attributes:

1. ID: Customer ID
2. Age Customer's approximate age.
3. CustomerSince: Customer of the bank since. [unit is masked]
4. HighestSpend: Customer's highest spend so far in one transaction. [unit is masked]
5. ZipCode: Customer's zip code.
6. HiddenScore: A score associated to the customer which is masked by the bank as an IP.
7. MonthlyAverageSpend: Customer's monthly average spend so far. [unit is masked]
8. Level: A level associated to the customer which is masked by the bank as an IP.
9. Mortgage: Customer's mortgage. [unit is masked]
10. Security: Customer's security asset with the bank. [unit is masked]
11. FixedDepositAccount: Customer's fixed deposit account with the bank. [unit is masked]
12. InternetBanking: if the customer uses internet banking.
13. CreditCard: if the customer uses bank's credit card.
14. LoanOnCard: if the customer has a loan on credit card.

PROJECT OBJECTIVE: Build an ML model to perform focused marketing by predicting the potential customers who will convert using the historical dataset.

1. **Import and warehouse data:**
   - Import all the given datasets and explore shape and size of each.
   - Merge all datasets onto one and explore final shape and size.
2. **Data cleansing:**
   - Explore and if required correct the datatypes of each attribute
   - Explore for null values in the attributes and if required drop or impute values.
3. **Data analysis & visualization:**
   - Perform detailed statistical analysis on the data.
   - Perform a detailed univariate, bivariate and multivariate analysis with appropriate detailed comments after each analysis.
4. **Data pre-processing:**
   - Segregate predictors vs target attributes
   - Check for target balancing and fix it if found imbalanced.
   - Perform train-test split.
5. **Model training, testing and tuning:**
   - Design and train a Logistic regression and Naive Bayes classifiers.
   - Display the classification accuracies for train and test data.
   - Display and explain the classification report in detail.
   - Apply all the possible tuning techniques to train the best model for the given data. Select the final best trained model with your comments for selecting this model.
6. **Conclusion and improvisation:**
   - Write your conclusion on the results.
   - Detailed suggestions or improvements or on quality, quantity, variety, velocity, veracity etc. on the data points collected by the bank to perform a better data analysis in future