**Digital Transformation in Banking Sector**

**Business Objective**

Bank XYZ has a growing customer base where the majority of them are liability customers (depositors) vs. borrowers (asset customers). The bank is interested in expanding the borrower’s base rapidly to bring in more business via loan interests.

A campaign that the bank ran in the last quarter showed an average single-digit conversion rate. In the last town hall, the marketing head mentioned that digital transformation being the core strength of the business strategy, how to devise effective campaigns with better target marketing to increase the conversion ratio to double-digit with same budget as per the last campaign.

As a data scientist, you are asked to develop a machine learning model to identify potential borrowers to support focused marketing.

**Data Description**

The dataset has 2 CSV files,

* Data1 - 5000 rows and 8 columns
* Data 2 - 5000 rows and 7 columns

The data consists of the following attributes:

1. ID: Customer ID

2. Age Customer’s approximate age.

3. Customer Since: Customer of the bank since.

4. Highest Spend: Customer’s highest spend so far in one transaction.

5. Zip Code: Customer’s zip code.

6. Hidden Score: A score associated to the customer which is masked by the bank as an IP.

7. Monthly Average Spend: Customer’s monthly average spend so far.

8. Level: A level associated to the customer which is masked by the bank as an IP.

9. Mortgage: Customer’s mortgage.

10. Security: Customer’s security asset with the bank.

11. Fixed Deposit Account: Customer’s fixed deposit account with the bank.

12. Internet Banking: if the customer uses internet banking.

13. Credit Card: if the customer uses bank’s credit card.

14. Loan On Card: if the customer has a loan on credit card

**Aim**

Build a machine learning model to perform focused digital marketing by predicting the potential customers who will convert from liability customers to asset customers.

**Tech stack**

* Language - Python
* Libraries – numpy, pandas, matplotlib, seaborn, sklearn, pickle, imblearn

**Approach**

1. Importing the required libraries and reading the dataset.

* Merging of the two datasets
* Understanding the dataset

2. Exploratory Data Analysis (EDA) –

* Data Visualization

3. Feature Engineering

* Dropping of unwanted columns
* Removal of null values
* Checking for multi-collinearity and removal of highly correlated features

4. Model Building

* Performing train test split
* Logistic Regression Model
* Weighted Logistic Regression Model
* Naive Bayes Model
* Support Vector Machine Model
* Decision Tree Classifier
* Random Forest Classifier

5. Model Validation

* Accuracy score
* Confusion matrix
* Area Under Curve (AUC)
* Recall score
* Precision score
* F1-score

6. Handling the unbalanced data using imblearn.

7. Hyper parameter Tuning (GridSearchCV)

* For Support Vector Machine Model

8. Creating the final model and making predictions

9. Save the model with the highest accuracy in the form of a pickle file.