

**PROJECT TITLE**: Bank Risk Controller Systems

**DOMAIN**: BANKING

SUBMITTED BY: SUBHIKSHA G R (MDTM28)

SUBMITTED ON: 25.12.2024

**EXECUTIVE SUMMARY:**

* The Bank Risk Controller System leverages advanced Machine Learning (ML) techniques to evaluate and predict financial risk within a banking environment.
* By analyzing historical transaction data, customer profiles, and external factors, the system aims to identify potential high-risk customers and prevent fraud.
* This project uses various ML models, including Random Forest, XGBoost, and Logistic Regression, to build a predictive framework for assessing the likelihood of defaults or fraudulent activity.
* The results of this analysis demonstrate significant improvements in risk prediction accuracy compared to traditional methods, providing a valuable tool for proactive risk management.

**PROBLEM STATEMENT EXPLAINATION:**

**Banks and financial institutions face a big challenge:**

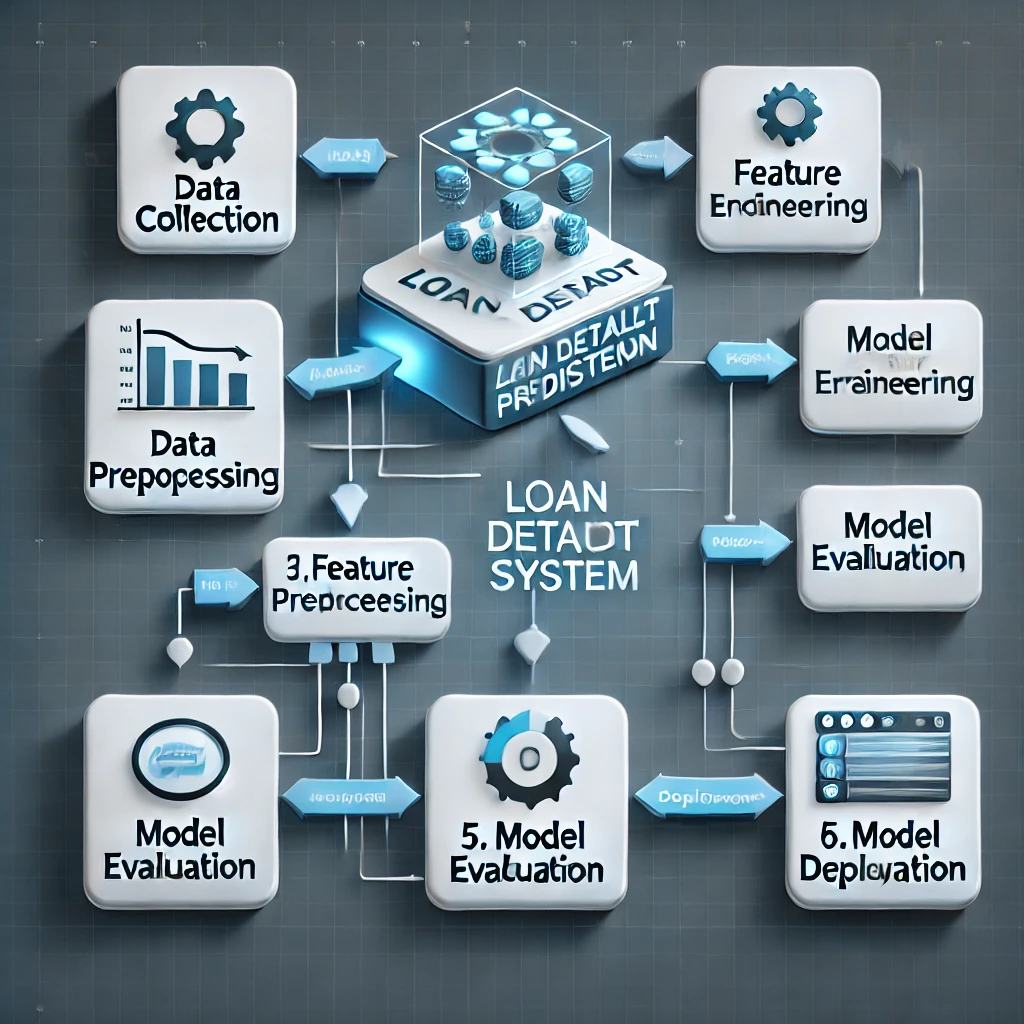
How to decide whether a customer applying for a loan will pay it back or fail to do so (default). If customers default on their loans, it can lead to financial losses for the bank. Predicting which customers are likely to default is important for making smarter loan approval decisions and reducing risks.

The goal of this project is to create a machine learning model that analyzes past loan data and predicts whether a new customer will default or not. The prediction is based on information like the customer’s income, loan amount, credit history, and other details. The target column in the dataset is called **“TARGET”**, where:

* **1** means the customer defaulted (did not repay the loan).
* **0** means the customer successfully repaid the loan.

**FLOWCHART:**

* Data Collection
* Data Preprocessing
* Feature Engineering
* Model Training
* Model Evaluation
* Deployment

****

**APPROACH:**

DATA COLLECTION

The dataset used for this analysis contains comprehensive information related to historical loan records and customer details. It is a robust dataset suitable for building predictive models.

|  |  |  |
| --- | --- | --- |
| S.NO | ASPECT | DETAILS |
| 1 | Total Records | 14,13,701 |
| 2 | Total Features | 158 |
| 3 | Data Format | CSV |
| 4 | Usage Purpose | Predict loan default probabilities |
| 5 | Data Source | GUVI |
| 6 | Nature of the Dataset | Mix of Continuous and Categorical variables relevant to the financial domain |

DATA PRE-PROCESSING

TARGET COLUMN ANALYZING

* The data has

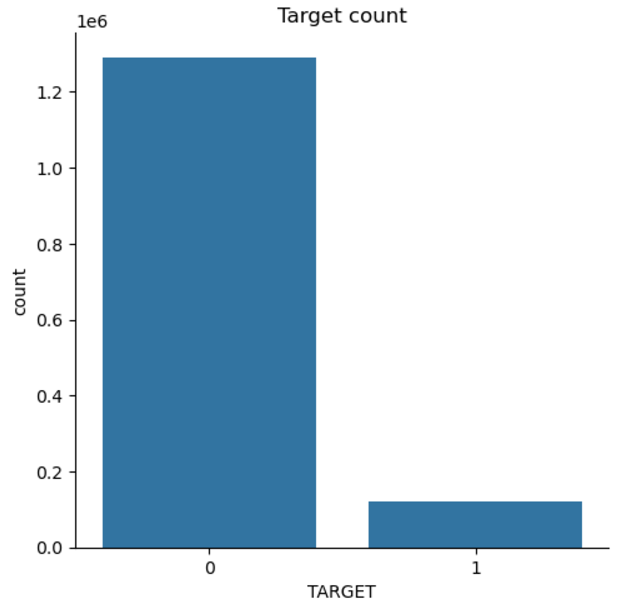
Target = 0 for people who can repay the loan and

Target = 1 for people who cannot repay loan. The data is imbalanced.

TARGET

0 1291341

1 122360



CASE SCENARIOS

(Scenario 1) - If the model has predicted client will repay loan but actually, he has defaulted.

(Scenario 2) - If the model has predicted client will default but he can actually pay loan back.

* But the loss will be much more in Scenario 1 i.e. If the model has predicted client will repay loan but actually, he has defaulted.
* Accuracy cannot be used for imbalanced data, hence not helpful in our case. Precision will tell of all the points that are predicted to be positive how many of them are actually positive. Precision doesn't consider, the points that were actually positive and are predicted positive.
* Recall is important since it will tell of all the points that are actually positive how many of them are predicted positive. So recall is important to identify the Scenario 1. If recall is high means defaulters are correctly predicted as defaulters
* F1-score is geometric mean of precision and recall and precision and is high if recall and precision are high. Since we have no use of precision here, F1-score or micro F1-score or macro F1-score is not important.
* In ROC-AUC score we have True Positive Rate and False Positive Rate; True Positive Rate is same as Recall so Scenario 1 gets covered. False Positive Rate is out of all the points that are actually negative how many are predicted as positive Scenario 2.
* Hence ROC-AUC score is an important metric since it covers both scenarios where Home Credit can suffer loss. ROC-AUC score should be greater than 0.5 which means your model is doing something right.

 Handling missing values, outliers, and categorical variables.

 Normalization and scaling techniques.

