**Project Report**

**Credit Card Behaviour Score Prediction Using Classification and Risk-Based Techniques**

Name- Lakshya Gupta

Enrollment number-24410015

Branch-Geological Technology

1. **Overview :**

Bank A wants to create a Behaviour Score, used for proactively identifying credit card customers that are likely to default the following month. We developed and assessed classification models to predict the likelihood of a default using anonymized behavioural and demographic data.

This report captures:

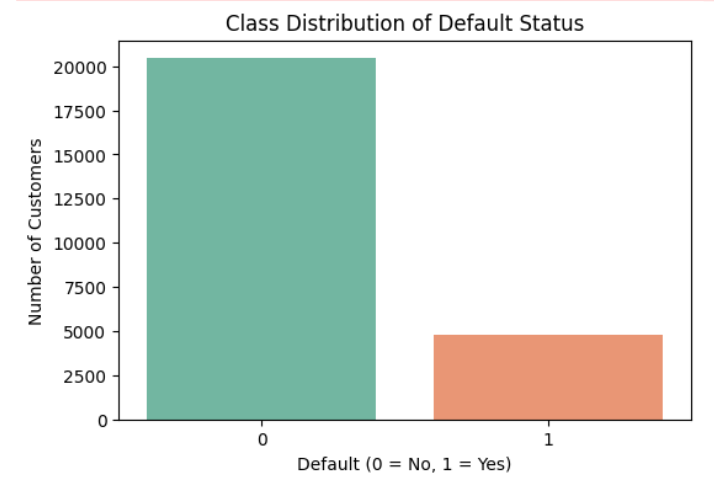
i.Exploratory data analysis (EDA) and finance orientated interpretation of customer behaviour

ii.Feature engineering and class imbalance handling

iii.Classification model development and evaluation, accounting for F2 score

iv.Classification decision making business implications

1. **Data Understanding & Preprocessing**

****

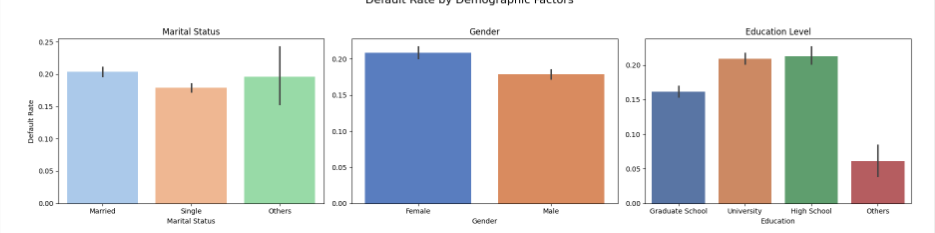
The dataset is imbalanced, with a significantly higher number of non-defaulters (0) compared to defaulters (1), highlighting the need for metrics like F2 score and possibly resampling techniques.

**2.1 Dataset Description**

* Training Data: ~25,000 rows with features like LIMIT\_BAL, AGE, SEX, EDUCATION, MARRIAGE, PAY\_m, BILL\_AMT\_m, PAY\_AMT\_m, etc.
* Validation Data: ~5,000 rows, identical features but without the target variable (next\_month\_default)
* Target Variable: next\_month\_default (1 = default, 0 = no default)

**2.2 Demographics**

* Gender: Female more likely to default (1 = Male)
* Marital Status: Married had a slightly higher default rate
* Education: University and High School groups show higher default rates (~21%)



**2.3 Behavioral Trends**

* Payment Status (PAY\_m):
  + Strongest predictor of default
  + Defaults rise with higher delays (PAY ≥ 1)
  + Customers with repeated negative payment statuses are lower risk
* Bill Amount (BILL\_AMT\_m) & Payment Amount (PAY\_AMT\_m):
  + Defaulting customers often had **high bill amounts and low payments**
  + Inconsistency in monthly payments is a major red flag

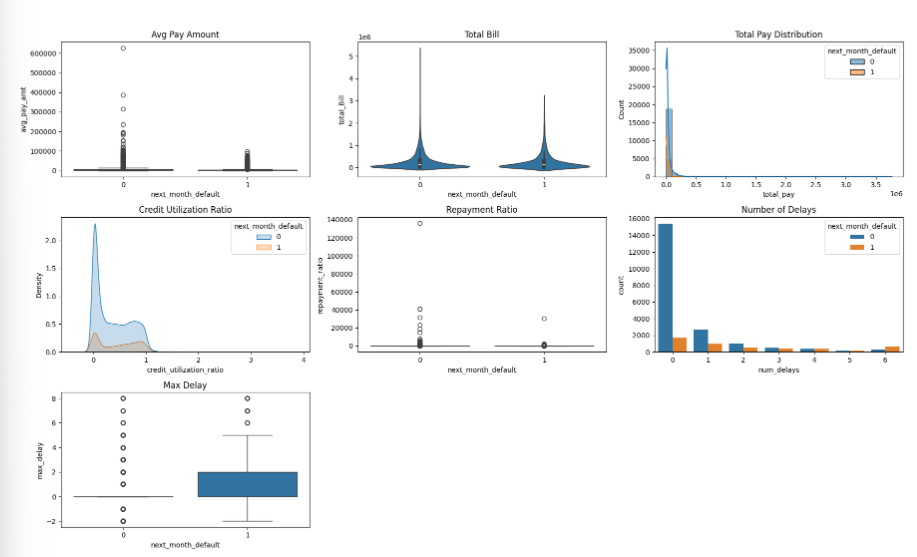
**2.4 Financial Ratios**

* PAY\_TO\_BILL\_ratio:
  + Strong discriminator — lower ratio indicates financial stress
* Utilization Ratio = Total Bill / Credit Limit
  + High utilization is correlated with higher default risk

**3. Feature Engineering**

**3.1 Financial Features Created**

* AVG\_Bill\_amt: Mean of BILL\_AMT1-6
* PAY\_TO\_BILL\_ratio: Total PAY\_AMT / Total BILL\_AMT
* Credit\_Utilization: AVG\_Bill\_amt / LIMIT\_BAL
* Delinquency\_Streak: Count of months with PAY\_m ≥ 1
* Consistent\_Repayment: Count of months with PAY\_m in {-1, 0}



These features added financial interpretability and improved model performance.

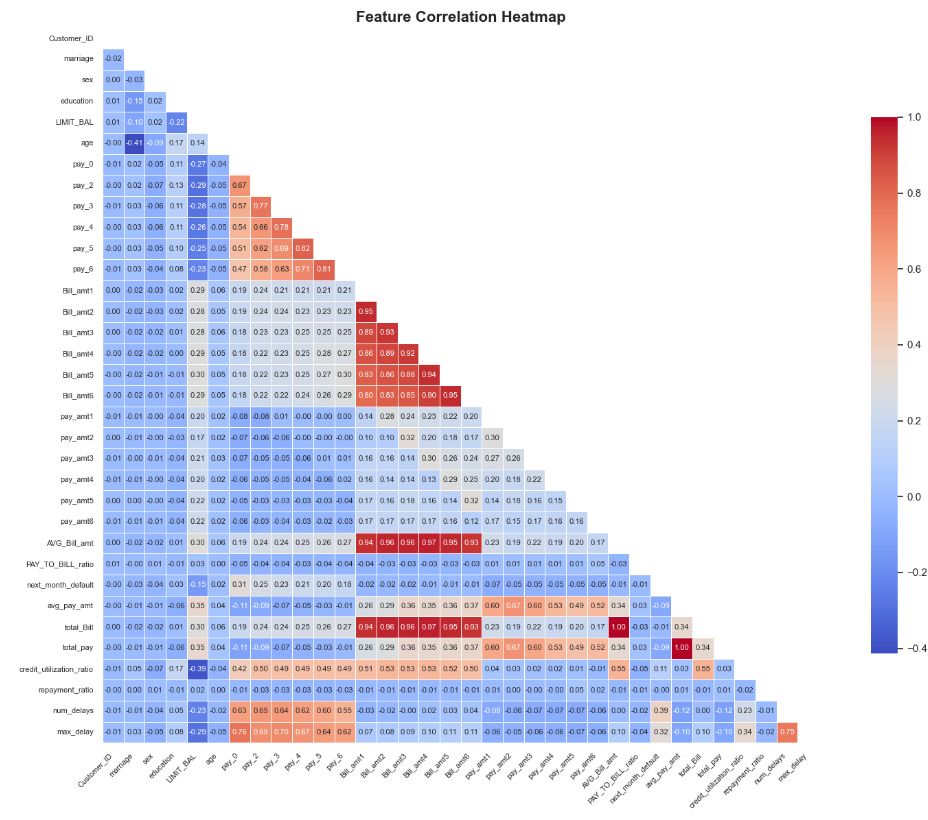
**3.2 Feature Correlation Analysis**

**Key Insights**:

1.High positive correlation was observed among billing features (Bill\_amt1 to Bill\_amt6), which may introduce redundancy.

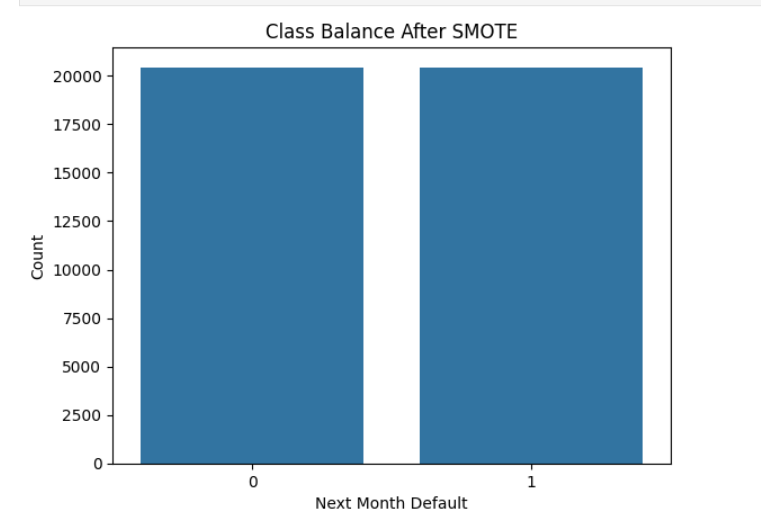
2.The target variable next\_month\_default shows moderate positive correlation with num\_delays, max\_delay, and repayment\_ratio.

3.Features with high correlation were reviewed for potential dimensionality reduction and model simplification.



**4. Handling Class Imbalance –**

1.Used SMOTE to oversample the minority class .



2.Compared it with using class weights into models like Logistic Regression and XGBoost 3.SMOTE showed improved results in terms of validation set F2-scores

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**5.. Model Building and Evaluation**

**5.1 Evaluation Metrics**

* **F2-score** prioritized over F1-score to give higher importance to recall (i.e., catching more defaulters)
* Also tracked Accuracy, Precision, Recall, ROC-AUC

**5.2 Model Selection**

**i. All the models used are evaluated on following metrices:**

1.Accuracy

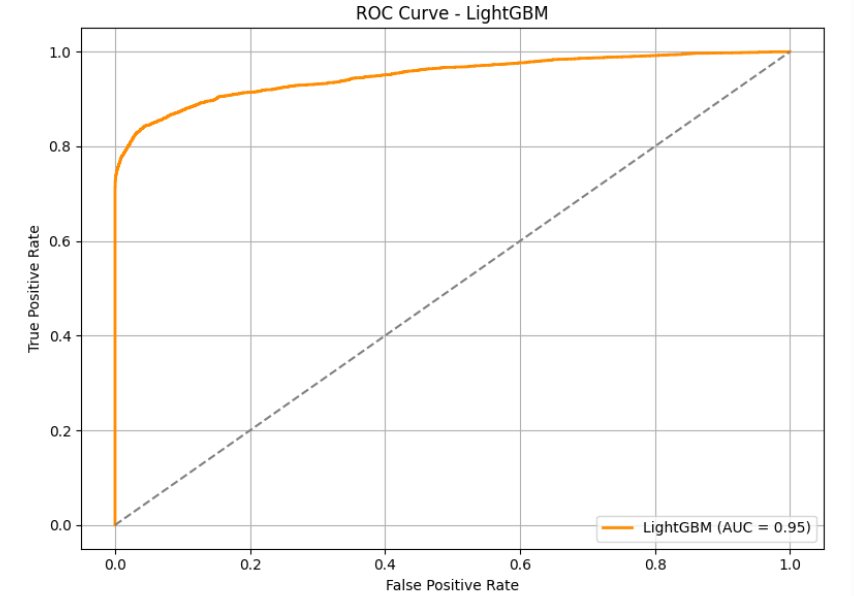
2.Precision

3.Recall

4.AUC-ROC

5.F2-Score

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
| **Model** | **Accuracy** | **F2 Score** | **AUC-ROC** | **Recall (Class 1)** | **Precision (Class 1)** | **Comments** | |
| **Logistic Regression** | 0.748 | 0.858 | 0.907 | 0.92 | 0.69 | Good AUC and F2, but recall tradeoff for Class 0 | |
| **Decision Tree** | 0.845 | 0.850 | 0.845 | 0.85 | 0.84 | Balanced results, interpretable model | |
| **XGBoost** | 0.842 | 0.890 | 0.947 | 0.85 | 0.90 | Strong F2 score and AUC, slightly lower recall | |
| **LightGBM** | **0.899** | **0.861** | **0.946** | **0.89** | 0.95 | **Best overall performance**, high recall and precision | |



After applying all model , the best model comes out to be LightGBM

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**6. Evaluation Metrics & Threshold Selection**

Why use F2 Score?

1.In the context of credit risk recall is much more important than precision.

2.Missing a defaulter (false negative) can cause a significant monetary loss, while right flagging a non-defaulter (false positive) is typically only a minor inconvenience.

3.To reflect this standard, we will use the F2-score, which makes recall more important than in F1-score.

4.The final classification threshold was selected at the point that optimized the F2-score on the validation dataset.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**7. Business Implications**

7.1 Understanding How Model Errors Relate to Business-

**1.False Positives (Predicted Default but not a Default)-**

i. Can lead to denial or limitation of credit to legitimate customers.

ii. Can lead to customer unhappiness and eventually impact customer retention.

**2.False Negatives (Predicted No Default but is a Default)**

i. Can lead to significant financial harm because we are failing to detect customers that , may default.

ii. Can eventually result in unlimited credit losses properly minimizing risk will be , directly associated to keeping customers well informed.

The threshold we selected was a business aware balancing act – minimizing false negatives but accepting a measurable level of false positives to balance risk exposure with customer experience.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**9. Predictions on Validation Dataset**

* Predictions generated using tuned XGBoost model and threshold = 0.38
* Submission file: submission\_<EnrollmentNumber>.csv

**10. Key Learnings**

* Financial domain knowledge is essential for meaningful feature engineering
* Handling imbalance and choosing the right metric can dramatically shift model performance
* Simple models (like logistic regression) with strong features can perform competitively
* Threshold tuning based on business needs can significantly affect decision-making