Ishu Swarnkar

23324007

BS-MS Physics(BP1)

1. Introduction

Credit risk management is an essential part of a bank's operations. It has a direct impact on profitability and financial stability. In this project, we aim to help the bank improve its credit risk framework by developing a predictive and interpretable Behaviour Score model. Our main goal is to identify credit card customers who are likely to miss their payments in the next billing cycle.

We use anonymized historical behavioral data from over 25,000 customers and apply machine learning classification techniques to create a forward-looking model that predicts defaults. By identifying potential defaulters early, the bank can adjust credit limits, activate early warning systems, and improve its decision-making process related to risk. This project connects data science with financial strategy, providing a model that helps in managing credit exposure more effectively.

2. EDA ANALYSIS

We did univariate and bivariate analysis of many features of the training dataset such as the analysis between the defaulters and the customers of different sex.

NOTE:- I changed the name and numerical entries of various columns into different names, they are clearly mentioned in my code file(ipynb)(made in Google Colab). I did so to simplify my EDA analysis while plotting graphs(though it was not required). However, for final training of models, I again reconsidered the old dataset to avoid confusion. I hope this helps as a clarification.

Key Observations from Analysis:

1. Customers with lower education levels, such as high school and others, show a higher default rate.

2. Single individuals are somewhat more likely to default than married ones.

3. Gender does not significantly impact default behavior. Most defaults occur among university graduates, mainly because they make up a larger portion of the dataset.

Conclusion:

Demographic factors, particularly education and marital status, offer helpful signals for assessing credit risk. However, they should be combined with behavioral and financial indicators for better prediction.

3. Feature Engineering

We featured various terms like delinquency streak, credit utilization ratio etc.

The terms which I engineered in this project are as follows:-

Delinquency Streak:- Delinquency Streaks are all about those months when payments are either late or completely missed. You can find this information in the columns labeled pay_0 through pay_6. It means if you see a value of 1 or more, it indicates that the payment is overdue by that many months. A streak is simply a series of consecutive pay_m values that are 1 or higher.

The Credit Utilization Ratio:- is an important factor in your credit score. It shows how much of your available credit you're actually using.

Repayment Consistency :- This aspect reveals how consistently a customer makes their payments and if those payments are adequate. It counts the number of months where pay_amt_m is greater than 0. It calculates the variance or standard deviation of pay_amt_m(less variation indicates more consistency). It tallies the number of on-time payments (where pay_m equals -1).

Overdue Months:- The total number of months a customer has missed making their full or minimum payment on time, falling behind on their bill by one month or more.

SIGNIFICANCE OF THESE TERMS:-

Significance of Overdue months: This metric reveals how frequently a customer misses their payment deadlines. A high number of overdue months indicates a tendency to delay payments, which is closely linked to their creditworthiness. It serves as a key sign of potential stress in their repayment habits. Banks often incorporate this information directly into their underwriting guidelines.

Significance of Credit Utilisation Ratio: This metric shows how much of their credit a customer is actually using. When the ratio is high (over 0.8), it indicates that the person is nearing their credit limit, which can lead to financial stress. On the other hand, a low ratio (around 0.3 to 0.5) is considered healthy. Banks often look at this as a clear sign of possible over-leveraging.

Significance of Delinquency Streak: This highlights ongoing financial struggles rather than just occasional setbacks. A prolonged period of delinquency signals a growing risk of default. It reveals a pattern of problematic behavior, not merely the frequency of issues.

Significance of Repayment Consistency: This measures how consistently a customer makes their payments. If there's a low variation in repayments or a high number of full and on-time payments, it indicates that the user is responsible. On the other hand, if repayments are inconsistent, it might point to cash flow problems or a lack of solid financial planning.

Further, we needed to go beyond basic accuracy and choose metrics that reflect real business risk. Here is the explanation:

- Recommended Metrics:

F2 Score:- Gives more weight to recall, which is crucial in credit risk. It's better to catch most defaulters, even with some false alarms.

Recall:- Measures how many actual defaulters were correctly identified. Missing a defaulter can lead to huge losses.

AUC-ROC :- Good for comparing models across thresholds. It shows how well the model separates defaulters from non-defaulters.

- Justification:

In credit risk, missing a defaulter (false negative) is more costly than wrongly flagging a safe customer (false positive). Therefore, we focus on the F2 score and recall, as they prioritize catching the most defaulters, which fits the bank's risk-averse nature.

2. Set a Classification Threshold

By default, classifiers use 0.5 as the decision threshold (that is, if probability is greater than 0.5, predict "default"). However, in banking, you might want to adjust that. We took it as 0.35(approximately)

- Business Implications:

False Positive:- Predicts default, but the customer would repay. Might reject a good customer, leading to lost revenue.

False Negative:- Predicts safe, but the customer defaults. High risk, resulting in direct financial loss.

Setting a lower threshold (for example, 0.3 to 0.4) if the bank is risk-averse. This approach aims to catch more defaulters (increase recall) while accepting more false positives.

We selected the F2 Score as the primary metric because it emphasizes recall. This focus is crucial in financial risk, where missing a defaulter is riskier than rejecting a good applicant. Additionally, we adjusted the classification threshold to fit the bank's risk appetite. This change improves our model's ability to flag high-risk profiles early, even if it results in a few more false positives.

4. Handling Class Imbalance

We applied SMOTE (Synthetic Minority Oversampling Technique) and class weighting to balance the dataset before training. This helped the model learn patterns from both defaulting and non-defaulting customers equally.

5. Modeling and Evaluation

Models evaluated:
- Logistic Regression: 70 % accuracy
- Decision Tree: 80% accuracy
- Random Forest: 88% accuracy
- KNN: 80% accuracy
- XGBoost: 88% accuracy
- LightGBM: 88% accuracy (best performer)

6. Threshold Tuning

Default threshold of 0.5 was suboptimal. Best results were achieved at threshold = 0.35:

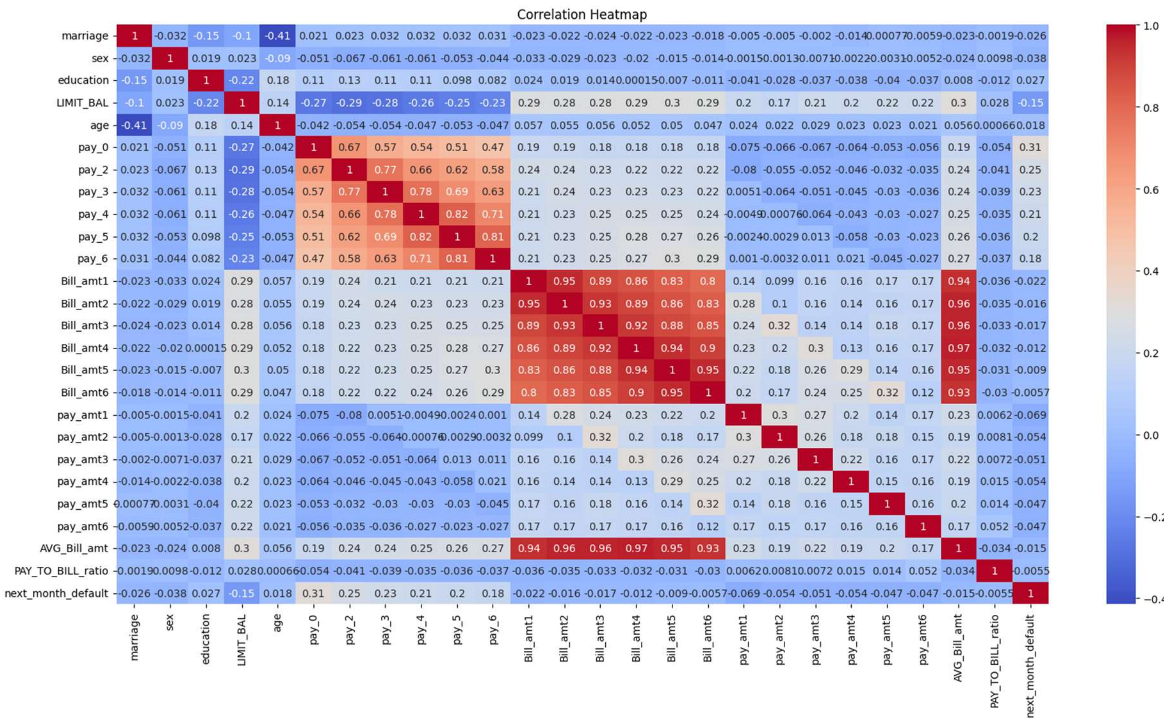
8. Final Remarks

Final Model: The best model is RandomForestClassifier after considering threshold as well.
Threshold: 0.35

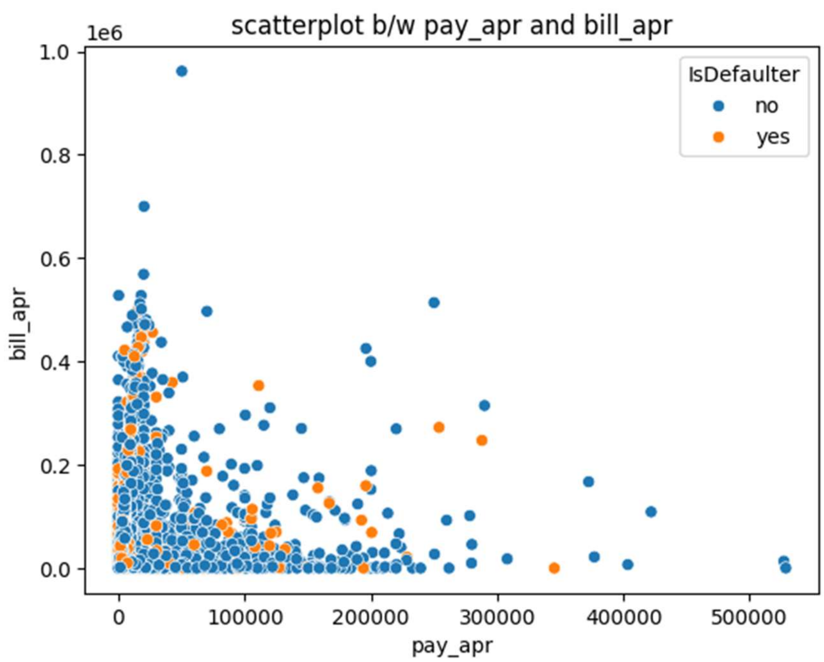
It was balanced with SMOTE and scaled with StandardScaler.
It is suitable to deploy for practical applications.

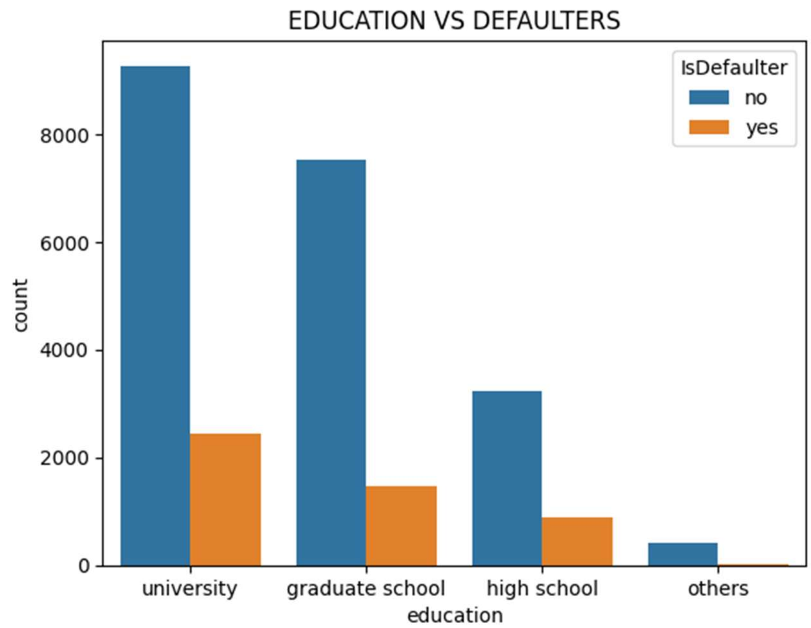
9. Visualizations

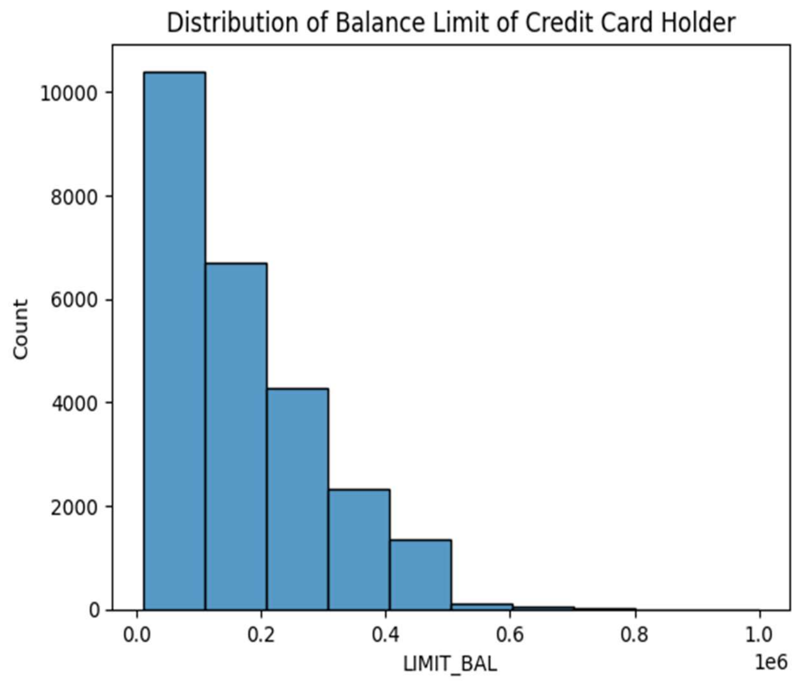
Below are some key graphs from the EDA analysis:



CORRELATION HEATMAP MATRIX







These are some of the plots from the bivariate and multivariate analysis done on different columns of the training dataset. (Note:- All the graphs made are present in the notebook itself so it is not feasible to include all in this report; I hope this helps.)

10. Conclusion

After thoroughly evaluating various machine learning models and their performance metrics, we found that RandomForestClassifier stood out as the top classifier. It showcased the best accuracy, AUC-ROC, and a solid balance between precision and recall. We further fine-tuned the model by adjusting the decision threshold to 0.35, which significantly enhanced its ability to identify defaulters while keeping false positives in check. Hence, we decided that the final model for predicting credit card defaults will be RandomForestClassifier with a threshold of 0.35, as it delivers the most dependable and understandable performance for real-world risk assessments.

Finally, we were able to generate a csv file which helped in predicting next_month_default in the validation dataset.

This project brilliantly merges financial principles with machine learning, creating a practical tool for predicting credit card defaults. It's crafted to help banks reduce risk, manage credit issuance, and enhance their financial decision-making processes.

(NOTE:- All the metrics and other observations are mentioned in the Colab notebook.)

Thank You!