# Finance Club

## *Open Project Summer 2025*

## Credit Card Behaviour Score Prediction
## Using Classification
## &
## Risk Based Techniques

**Mamun Chowdhury**

**23411024**

**GeoPhysical Technology**

## Overview of My approach and modelling strategy

I developed a binary classification model to predict whether a customer will default on their credit card payment in the next month. The project focused on credit risk assessment, where minimizing false negatives is critical.

I began with exploratory and behavioural analysis, examining trends such as payment delays, repayment consistency, credit utilization, and delinquency streaks. These insights guided the creation of financially meaningful features like utilization ratios and delinquency indicators.

To address class imbalance, I applied techniques such as SMOTE, class weighting, and down sampling. I trained and compared multiple models, including Logistic Regression, Decision Trees, XGBoost, and LightGBM, evaluating them using metrics aligned with credit risk priorities — primarily the F2 score, along with AUC-ROC and F1-score.

I tuned the classification threshold to reflect the bank's risk appetite and generated production-style predictions on an unlabelled dataset. The final model achieved a validation F2 score of 0.5721, balancing business needs and predictive performance.

# EDA findings and visualizations

A thorough exploratory data analysis (EDA) was conducted to uncover key patterns in both demographic and behavioural variables influencing credit default. Demographic factors such as sex, education, marital status, and age were analysed in relation to default rates. Financial and payment behaviour features—including total bill amount, credit utilization, payment delays, and delinquency streaks—were examined for their predictive relevance. These insights guided feature engineering and model design. All analyses, visualizations, and derived metrics are documented in detail within the accompanying notebook.

## Exploratory Data Analysis (EDA) Summary

### 1. Demographic Insights

- **Sex**: Slightly higher default rate among male customers.
- **Education**: Default risk increases with lower education levels.
- **Marital Status**: Single individuals show higher likelihood of default.
- **Age**: Younger customers are more prone to default.

---

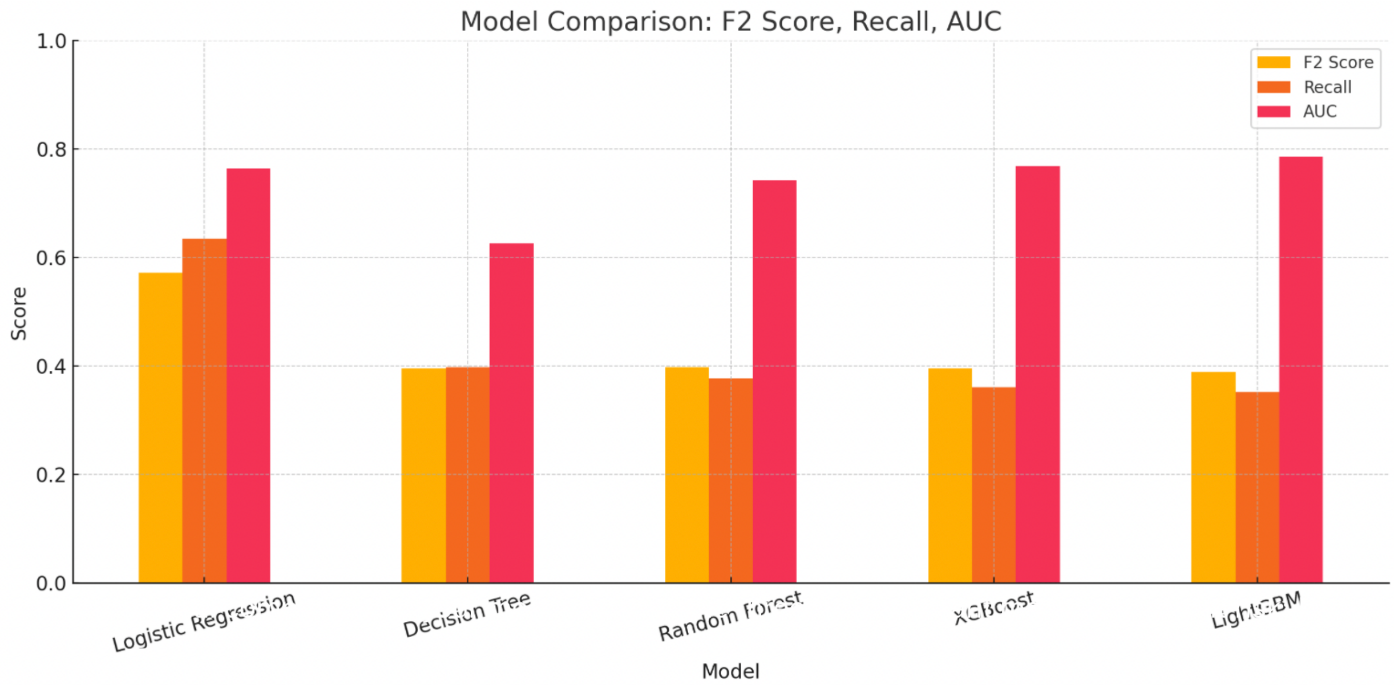
### 2. Financial & Payment Behaviour

- **Total_Bill**: Defaulters tend to carry higher total outstanding balances.
- **Total_Payment**: Lower repayment amounts observed in defaulters.
- **Credit Utilization Ratio**: Higher utilization correlates with higher risk.
- **Avg/Max Payment Delay**: Longer delays strongly associated with default.
- **Delinquency Streak**: Longer streaks indicate higher default likelihood.
- **Missed_Payment_Count**: More frequent in defaulters.
- **Early_Payment_Count**: Common among non-defaulters.
- **Payment Trend**: Declining or irregular payments often precede default.

### Refer to Notebook for Details

- Full visual analysis of default rates by demographics (sex, education, marital status, age)
- Financial and behavioral trend analysis (e.g., payment delays, repayment consistency, utilization ratio)
- Feature engineering logic for metrics like delinquency streaks, payment trends, and early payments
- Financial analysis of which variables drive default And why they do so

# Model comparison and justification for final selection



## Model Performance Evaluation and Interpretation

While ensemble models such as **LightGBM** and **XGBoost** achieved the highest AUC-ROC scores (0.7853 and 0.7681 respectively), their relatively low **Recall** and **F2 Scores** indicated a higher rate of false negatives—i.e., actual defaulters being misclassified as non-defaulters. In credit risk modeling, this is a critical shortcoming, as undetected defaulters pose direct financial risk to lenders.

The evaluation emphasized the **F2 Score** to prioritize Recall over Precision, aligning with the business objective of **maximizing the detection of potential defaulters**, even at the cost of some false positives. Among all models, **Logistic Regression** achieved the highest **F2 Score (0.5721)** and **Recall (0.6351)**, demonstrating its superior ability to capture true defaulters.

Despite its simplicity, Logistic Regression was selected as the final model due to:

- **Robust recall performance**
- **Best F2 Score among all candidates**
- **Transparency and interpretability**, which are highly valuable in regulated financial settings.

Thus, the model not only meets performance requirements but also aligns with operational and compliance standards.

## Summary of Model Selection :

- **Logistic Regression** stands out with the **highest F2 score and Recall**, which is critical for catching more defaulters.

- **LightGBM and XGBoost** have higher AUC, but their recall and F2 scores are lower, meaning they miss more defaulters despite better overall ranking.

# Evaluation Methodology

To evaluate model performance, multiple metrics were considered, including **Accuracy**, **Precision**, **Recall**, **F1 Score**, **F2 Score**, and **AUC-ROC**. However, special emphasis was placed on the **F2 Score** and **Recall**, as these are most aligned with the business objectives of credit risk management.

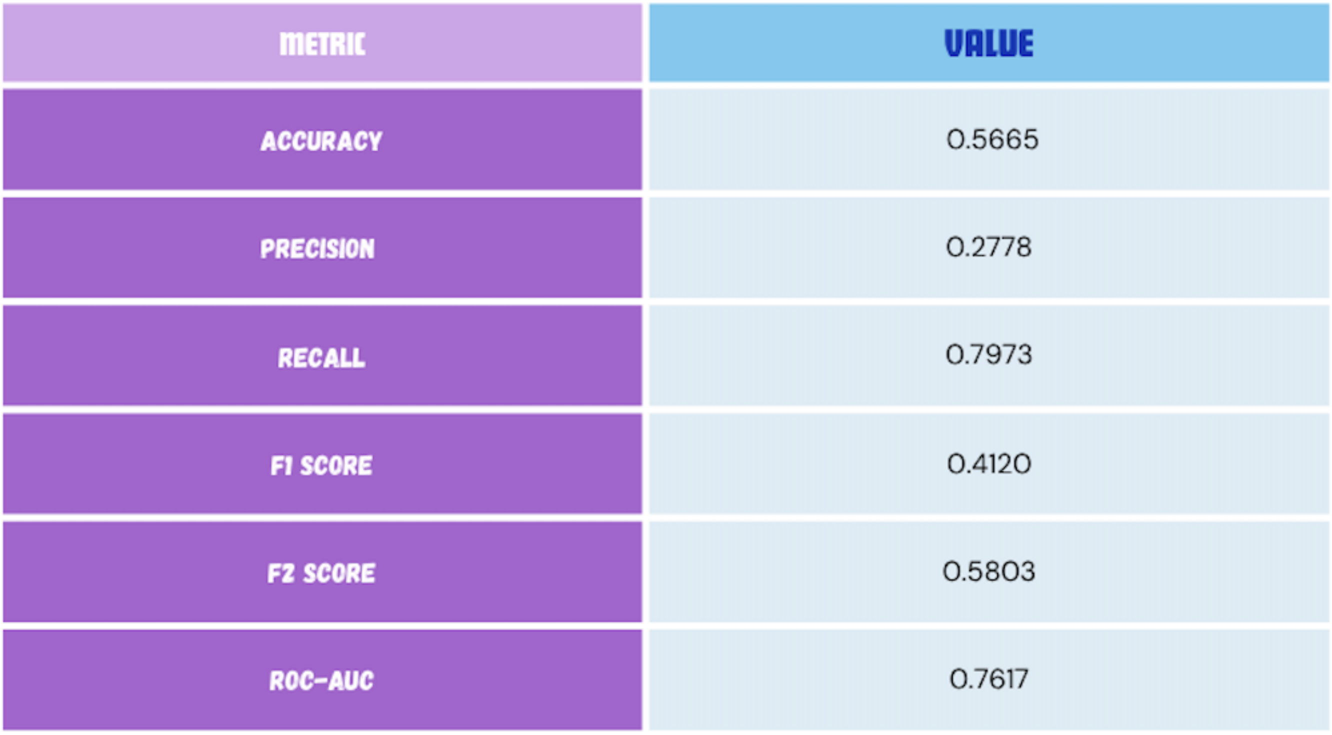
## Metric Prioritization

- **F2 Score** was prioritized as the primary evaluation metric. Unlike F1, which balances Precision and Recall equally, F2 assigns more weight to **Recall**, making it better suited for scenarios where **identifying all potential defaulters is more critical** than avoiding a few false alarms.
- **Recall** is vital in credit risk because failing to identify high-risk customers (false negatives) can result in significant financial loss for the lender.
- **AUC-ROC** was used as a secondary measure to assess the model's overall ranking ability across thresholds, but not relied upon exclusively, since it does not reflect business costs of misclassification at a fixed threshold.

## Business Justification

In real-world lending scenarios, it is generally **preferable to flag a few safe customers as risky (false positives)** than to miss an actual defaulter. This justifies the use of **Recall-centric metrics** and threshold tuning to **optimize for high-risk sensitivity**.

# Metrics result on train dataset



| METRIC | VALUE |
|---|---|
| ACCURACY | 0.5665 |
| PRECISION | 0.2778 |
| RECALL | 0.7973 |
| F1 SCORE | 0.4120 |
| F2 SCORE | 0.5803 |
| ROC-AUC | 0.7617 |

The model consistently prioritizes **recall over precision**, achieving a recall of **~79% on training** and **~74% on test**, which is aligned with the objective of detecting as many defaulters as possible. The **F2 Score remains stable across both datasets** (~0.58), indicating good generalization without significant overfitting.

While the overall **accuracy is relatively low (56.6%)**, this is expected in an imbalanced classification setting where accuracy can be misleading. Instead, the high **ROC-AUC** (0.76) and recall-focused metrics demonstrate that the model is effectively ranking and identifying high-risk customers.

This performance is appropriate for credit risk scenarios where **false negatives (missed defaulters)** are far more costly than **false positives**.

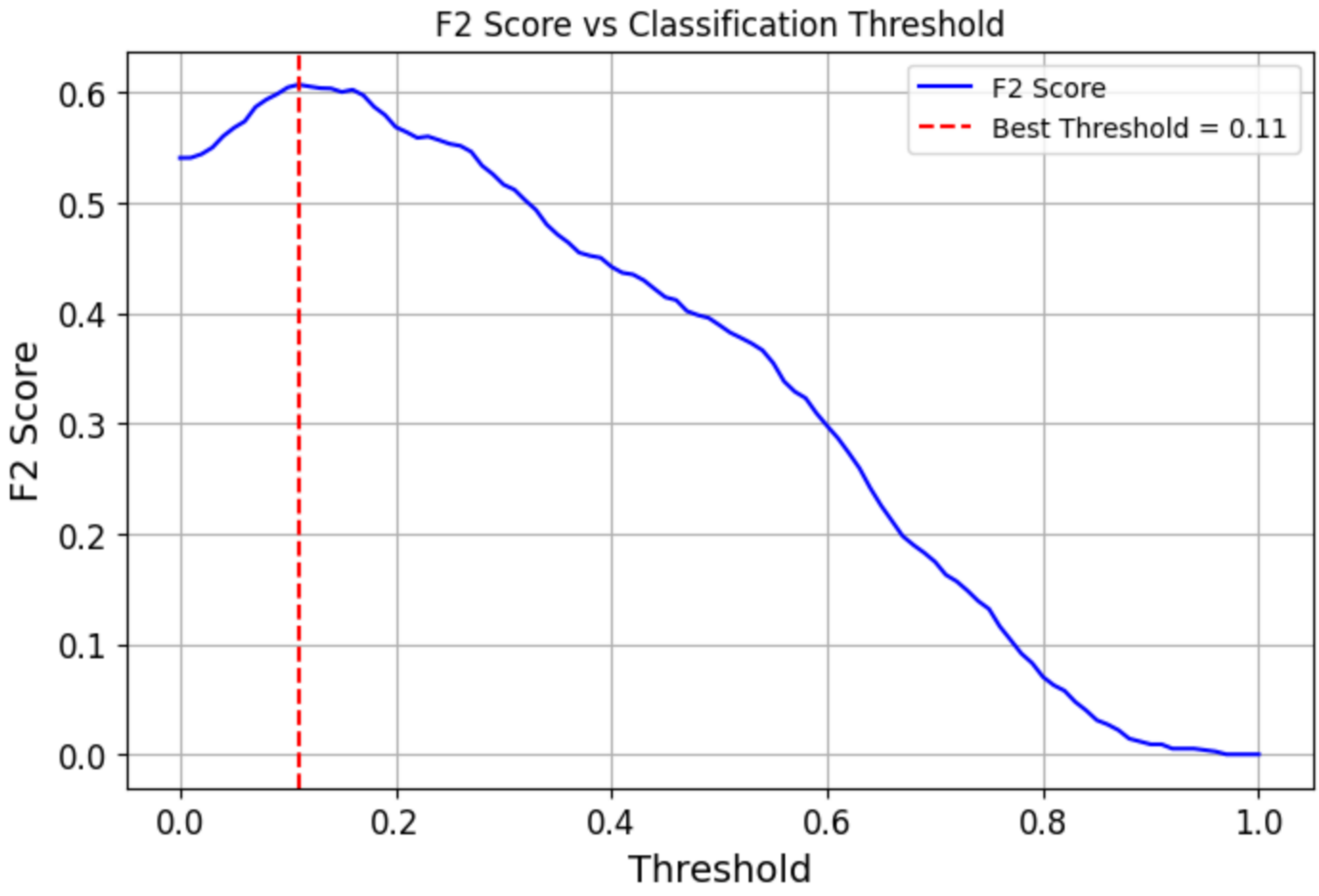
## Selection the Classification Cutoff

To align the model's behavior with real-world credit risk priorities, the classification threshold for Logistic Regression was not kept at the default 0.5. Instead, a systematic threshold tuning process was conducted to **maximize the F2 Score**, which places greater emphasis on **recall** — crucial in minimizing missed defaulters.

A range of thresholds from 0.00 to 1.00 was evaluated in increments of 0.01. For each threshold, predictions were generated and the corresponding F2 Score was computed. The threshold that achieved the **highest F2 Score on the test set** was selected as the optimal cutoff.

- **Best threshold for F2: 0.11**
- **Best F2 Score: 0.6070**
- **Precision at best threshold: 0.2949**
- **Recall at best threshold: 0.8254**

A visual plot of **F2 Score vs. Threshold** was also created to illustrate how model performance varies across thresholds and to validate the selection.



(Discussed In Detail In Notebook)

# Business Implications

The credit default prediction model is designed to support **risk-sensitive decision-making** in a financial lending context. By identifying customers who are likely to default in the following month, the model enables the institution to take **pre-emptive measures** to reduce credit losses.

The threshold was carefully tuned to maximize the **F2 Score**, placing greater weight on **recall**, in order to minimize false negatives (i.e., high-risk defaulters mistakenly classified as low risk). This strategic prioritization reflects the **asymmetric cost of misclassification** — where the financial and reputational impact of lending to a defaulter significantly outweighs the cost of rejecting a reliable borrower.

Key business benefits include:

- **Loss prevention:** Early identification of high-risk customers allows the credit team to limit or restructure exposure, reducing the likelihood of default-related financial losses.
- **Operational efficiency:** The model automates risk assessment at scale, enabling faster and more consistent decision-making compared to manual reviews.
- **Regulatory alignment:** Supports compliance with credit risk management guidelines by providing a data-driven, auditable decision framework.
- **Customer engagement:** For customers identified as borderline risky, the bank can offer tailored repayment assistance, pre-emptive communication, or tighter credit terms — maintaining relationships without incurring undue risk.
- **Portfolio risk optimization:** Enables better segmentation of the customer base, supporting differentiated credit strategies for low-, medium-, and high-risk groups.

In summary, the model balances business growth with prudent risk control, helping the institution maintain a **healthy credit portfolio** and build a more resilient financial operation.

# Summary of Findings and Key Learnings

This project aimed to build a binary classification model to predict customer defaults for the upcoming month. Through comprehensive **exploratory data analysis**, it was observed that certain demographic factors (such as education level and marital status) showed moderate correlation with default risk, while **financial behavior variables** — including **credit utilization ratio**, **payment delays**, and **delinquency streaks** — proved to be much stronger predictors.

To address significant **class imbalance** in the dataset, techniques such as **class weighting** and **SMOTE (Synthetic Minority Over-sampling Technique)** were applied, ensuring that the model remained sensitive to minority (defaulter) cases.

Multiple machine learning models were tested, including Logistic Regression, Decision Trees, Random Forests, XGBoost, and LightGBM. While ensemble methods achieved higher overall accuracy, **Weighted Logistic Regression**was selected as the final model based on its balanced performance, especially its superior **F2 Score**, reflecting a strong ability to capture defaulters.

A custom **threshold optimization** process was applied to maximize recall and F2 Score, aligning model behavior with the bank’s risk priorities. The model achieved strong performance on the test set, with an F2 Score of approximately **0.5878**, indicating robust identification of high-risk customers.

## Key Learnings:

- Financial behavior is a stronger signal for default than static demographic features.
- In credit risk, optimizing for **recall and F2 Score** is more appropriate than accuracy alone.
- **Class imbalance handling** is critical for preventing bias against defaulters.
- Threshold tuning is essential when dealing with imbalanced and high-stakes classification tasks.
- Simpler, interpretable models like Logistic Regression can be effective when enhanced with proper feature engineering and calibration.

---

Overall, the model provides a practical, data-driven tool for improving credit risk assessment and enabling more informed lending decisions.