# REPORT FOR CREDIT CARD DEFAULT PREDICTION

Soumya Gupta
23112101

## 1. Project Overview

The objective of this project is to build a predictive model that can identify customers who are likely to default on their credit card payments in the upcoming month. Early identification of high-risk customers is crucial for financial institutions to mitigate risk and optimize credit offerings.

- **Approach Summary**

Our workflow involved the following key steps:

1. **Data Exploration (EDA):**
   We began by analyzing the distribution of features, handling missing values, and studying correlations between variables and the target (`next_month_default`). Features like past payment status (`PAY_0` to `PAY_6`) and bill/payment history were strong indicators of risk.

**2. Feature Engineering:**
To better capture customer behavior and risk, we engineered several new features:

- **`avg_utilization_ratio`** – Average credit utilization over six months.
- **`num_delinquent_months`** – Count of months with delayed payments.
- **`longest_delinquency_streak`** – Longest continuous period of missed or late payments.

**3. Preprocessing and Class Imbalance Handling:**

- Applied one-hot encoding or label encoding where necessary.
- Scaled features for distance-based models.
- Addressed class imbalance using techniques like **SMOTE** to ensure the model does not favor the majority class.

**4. Model Training:**
We trained and compared five classification models:

- Logistic Regression – Baseline model for interpretability.
- Decision Tree – Captures basic non-linear patterns.
- Random Forest – Ensemble of trees for better generalization.
- XGBoost – Gradient boosting model, performed best overall.
- LightGBM – Fast, efficient boosting model.

**5. Hyperparameter Tuning:**
XGBoost's parameters (like `max_depth`, `learning_rate`, `n_estimators`, and `subsample`) were optimized using RandomizedSearchCV.

**6. Threshold Tuning:**
Instead of using the default 0.5 classification threshold, we tuned it to **maximize the F2 Score**, which places more emphasis on recall. This is especially important in credit risk to avoid **false negatives** (i.e., missing a defaulter). The final threshold after tuning came to be 0.2.

**7. Model Evaluation**

The model was evaluated using a comprehensive set of metrics to understand its predictive performance, especially in detecting defaulters.We used the following metrics:

- **Accuracy**: Overall correctness of the model.
- **Precision**: How many predicted defaulters were actually defaulters.
- **Recall**: How many actual defaulters were correctly identified.

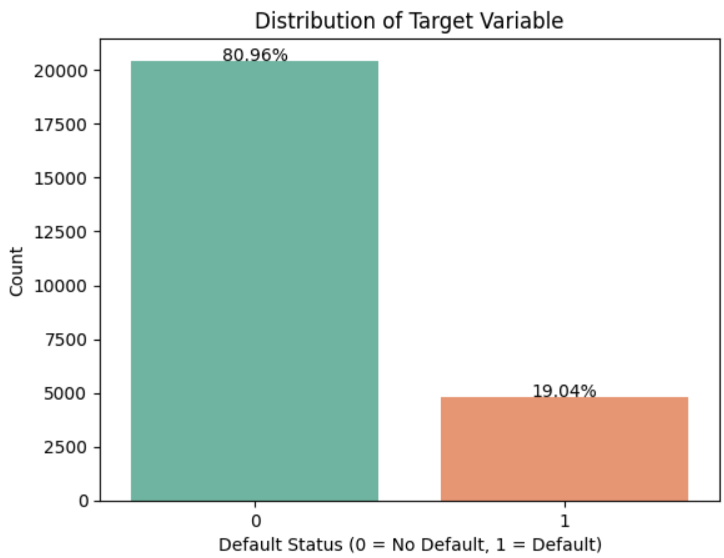
- **F1 Score**: Balance between precision and recall.
- **F2 Score**: Like F1, but gives more weight to **recall** — important for credit risk to avoid missing defaulters.
- **ROC-AUC Score**: Measures the model's ability to distinguish between defaulters and non-defaulters across thresholds.
- **Confusion Matrix**: Visual breakdown of true/false positives and negatives.

Among these, **F2 Score** was prioritized during threshold tuning to minimize **false negatives**, which are costlier in financial decision-making.

## 2. Exploratory Data Analysis (EDA) Findings and Visualizations

### Class Imbalance

The dataset is notably **imbalanced**, with a significantly lower proportion of defaulters (target = 1) compared to non-defaulters (target = 0).



## Variable Distributions

- **Demographic Variables:**
  - `Age` is right-skewed with most customers between **30 to 50 years**.
  - `LIMIT_BAL` (credit limit) shows high skewness; most customers have lower credit limits.

- **Financial Variables:**
  - All `BILL_AMT` and `PAY_AMT` features are **heavily right-skewed**, indicating that most customers deal with relatively low amounts, while a few outliers handle much larger sums.
  - `PAY_TO_BILL_ratio` is extremely dispersed with outliers, indicating inconsistent repayment strategies.

---

## Correlation Analysis

- **Overdue Payments (`pay_0` to `pay_6`)** show the **strongest positive correlation with default**, confirming that **delayed payments are a key risk indicator**.
- `LIMIT_BAL` and `AVG_Bill_amt` have **weak to moderate negative correlations** with default, suggesting that **customers with lower limits and higher usage may be riskier**.
- Strong multicollinearity is observed among `BILL_AMT` and `PAY_AMT` features, reflecting consistency in financial behavior over months.

---

## Demographic Patterns

- **Education + Gender:** University and high school-educated males show **slightly higher default rates**, possibly due to more financial responsibilities or risks.
- **Marital Status + Education:** Individuals from the "Other" marital status group, particularly with lower education, **exhibit the highest default rates**.
- Overall, demographic variables have **visible trends but lower predictive power** than financial indicators.

---

## 3. Financial Insights: Key Drivers of Default

### 1. Overdue Payment History

The strongest predictor of default is the **history of overdue payments**, as captured by `pay_0` to `pay_6`. Customers with a record of past payment delays are **far more likely to default again**, indicating behavioral risk.

### 2. Repayment Behavior and Consistency

- Defaulters tend to have **higher variability** in their `PAY_TO_BILL_ratio`, meaning their repayments are **inconsistent across months**.
- This instability suggests **poor financial planning or unpredictable cash flow**, both of which increase credit risk.

### 3. Credit Utilization & Repayment Insufficiency

- The scatter plots of **monthly payments vs. bills** indicate that many customers **underpay their bills**, even when amounts are large.
- This chronic underpayment may lead to **debt accumulation**, making default more likely in subsequent months.

# 4. Model Comparison And Final Prediction

## Metric Selection

In credit risk modeling, **the cost of missing a defaulter (false negative)** is much higher than incorrectly flagging a non-defaulter (false positive). Therefore, we prioritized metrics that emphasize recall.

The key metrics used were:**Recall,F2 Score**,**F1 Score,Accuracy**.

We evaluated all models using these metrics on both training and validation sets, **<u>but focused primarily on F2 Score and recall</u>** during threshold tuning and final selection.

```
[LightGBM] [Info] Start training from score -0.002691
|
Model Performance Comparison:

                     Accuracy  Precision    Recall  F1 Score  F2 Score  \
XGBoost              0.886742   0.922730  0.845499  0.882428  0.859893
LightGBM             0.884907   0.924003  0.840146  0.880082  0.855677
Random Forest        0.755015   0.795512  0.690024  0.739023  0.708823
Decision Tree        0.746208   0.780226  0.689294  0.731947  0.705745
Logistic Regression  0.691414   0.754735  0.572019  0.650796  0.601125

                      ROC-AUC
XGBoost              0.944971
LightGBM             0.944683
Random Forest        0.840513
Decision Tree        0.834951
Logistic Regression  0.756660

Best Model: XGBoost with F2 Score = 0.8599
```

## Final Model Selection

**XGBoost** was chosen as the final model because:

- It achieved the **<u>highest F2 Score (0.8599)</u>**, aligning with the project's objective to minimize false negatives.
- It also had the **highest ROC-AUC (0.9449)**, indicating excellent class discrimination.

- Compared to LightGBM, it showed slightly better recall and F2 Score, making it more suitable for credit risk prediction

## 5. Classification Cutoff Selectionl.

**Why Customize the Threshold?**

The default 0.5 cutoff may not be ideal for credit risk.

- **High threshold** (e.g., 0.7): Fewer false positives, but more defaulters missed.
- **Low threshold** (e.g., 0.3–0.5): Catches more defaulters (higher recall), but increases false positives.

Since **missing defaulters is costlier**, we tuned the threshold to **maximize recall and F2 Score**.

**Our Approach**

We conducted **threshold tuning** by varying the classification cutoff between **0.1 and 0.9** in increments ( 0.05). For each threshold, we computed the **F2 Score**, which emphasizes recall more than precision.
**Final Threshold Used**: `0.2`

## 6. Business Implications

The predictive model developed in this project has significant practical value for credit risk management:

- **Early Defaulter Detection**: By identifying high-risk customers before they default, banks and credit card companies can take preventive actions such as limiting credit, increasing interest rates, or initiating recovery strategies.

- **Reduced Financial Losses**: Accurately flagging likely defaulters helps minimize **bad debts** and improves overall financial health of the institution.
- **Smarter Credit Decisions**: The model supports **data-driven lending decisions**, reducing reliance on manual assessments and improving consistency.
- **Improved Risk-Reward Balance**: With better classification, companies can offer more credit to trustworthy customers while minimizing exposure to risky ones.

In summary, the model enhances both **operational efficiency** and **profitability**, helping financial institutions make safer and smarter lending decisions.

## 7. Summary of Findings and Key Learnings

Throughout the credit risk modeling project, we performed detailed exploratory data analysis, feature engineering, model development, and evaluation. Key insights and outcomes include:

- The dataset was **imbalanced**, with far fewer defaulters compared to non-defaulters, making **recall-focused metrics like F2-score** essential for evaluation.
- EDA revealed that **overdue payment features (PAY_0 to PAY_6)** had the strongest correlation with defaults, reinforcing the importance of repayment history.
- Features such as **credit utilization**, **repayment consistency**, and **bill-to-payment behavior** significantly influenced the likelihood of default.

- Demographic factors (e.g., **education, marital status**) also impacted default rates but were secondary to financial variables.

- Custom **feature engineering** (e.g., `avg_utilization_ratio`, `num_delinquent_months`, `longest_delinquency_streak`) improved model performance by capturing financial patterns better.
- A range of classification models was evaluated, including **Logistic Regression**, **Random Forest**, **XGBoost**, and **LightGBM**.

**Final Model & Evaluation:**

- ✅ **Best-performing model: XGBoost**
- ✅ **F2 Score after hyperparameter tuning: 0.867**
- ✅ **Selected classification threshold: 0.2**, optimized to **maximize recall** and minimize false negatives — aligning with the business goal of catching as many defaulters as possible.

```
Fitting 3 folds for each of 30 candidates, totalling 90 fits
Best Hyperparameters:
 {'subsample': 0.7, 'n_estimators': 300, 'min_child_weight': 3, 'max_depth': 8, 'learning_rate': 0.1, 'gamma': 0.2, 'colsample_bytree': 1.0}

Evaluation on Test Set:
Precision: 0.928042328042328
Recall: 0.8535279805352798
F1 Score: 0.8892268694550064
ROC-AUC: 0.9524561882227911
F2 Score: 0.8674579624134521
Accuracy: 0.8931017612524462
```