Credit Card Default Prediction Project Report

Name-Anushree Gupta
Enrolment No. 23115020

1. **Project Overview**

This project aims to predict the likelihood of a customer defaulting on their credit card payment in the upcoming month. Using a dataset containing credit behaviour and financial information, I explored multiple machine learning models to identify patterns and create a robust classifier.

Given the business context, F2-score was chosen as the key evaluation metric to emphasise recall more than precision. This is important because failing to catch defaulters can cost a financial institution heavily.

2. **Data Understanding & Preprocessing**

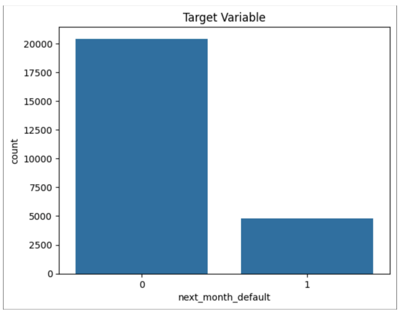
- The dataset consisted of demographic information, payment history, bill statements, and actual payment amounts over six months.
- Key columns included: 'LIMIT_BAL', 'sex', 'education', 'marriage', 'age', 'pay_1' to 'pay_6', 'Bill_amt1' to 'Bill_amt6', 'pay_amt1' to 'pay_amt6', Avg_Bill_amt and Pay_to_bill_ratio.

Cleaning Steps:

- Replaced invalid 'education' and 'marriage' values with grouped categories as some categories were not labelled in the problem statement and also were comparatively very less in number.
- Handled missing values in 'age' by imputing with the median to avoid the effect of outliers and preserve the central tendency in a simple
- Converted 'pay_0' to 'pay_1' for consistency.

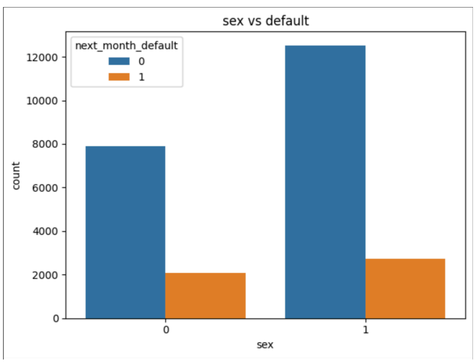
3. **Exploratory Data Analysis (EDA)**

Target Distribution: The dataset was imbalanced with fewer default cases.

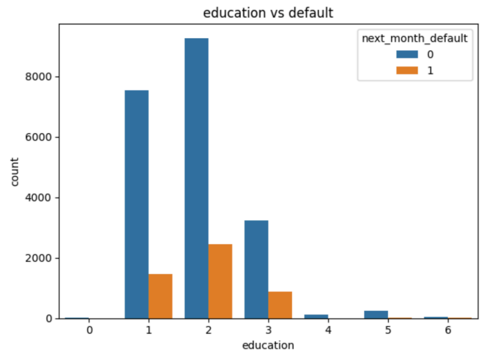


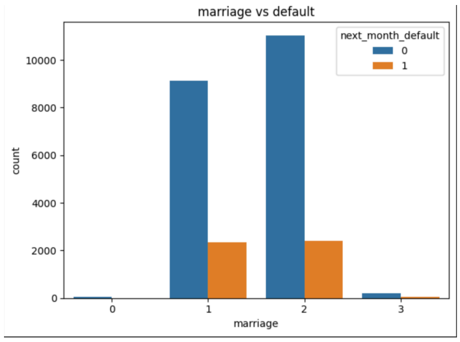
Categorical Variables:

* More males defaulted than females.



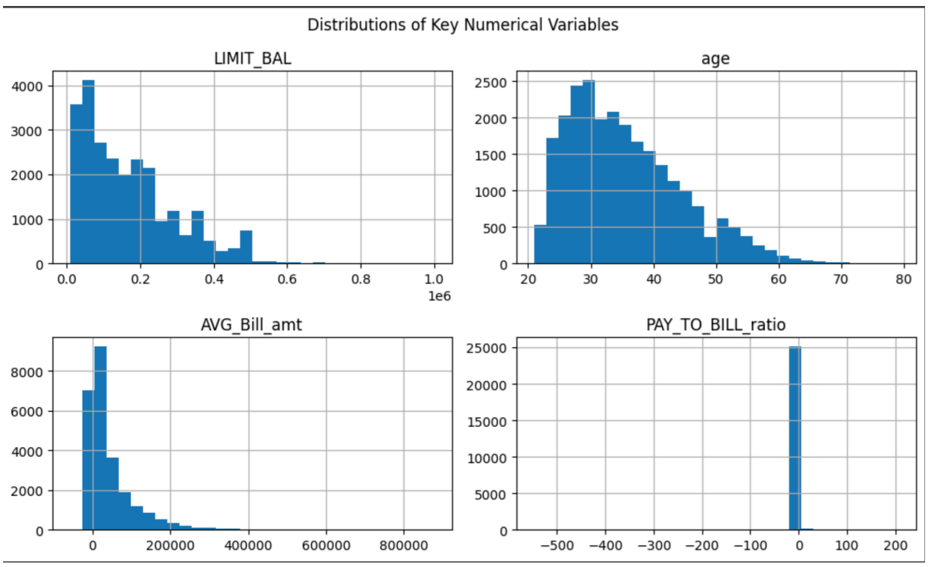
* Defaulting patterns were different across education and marital status.





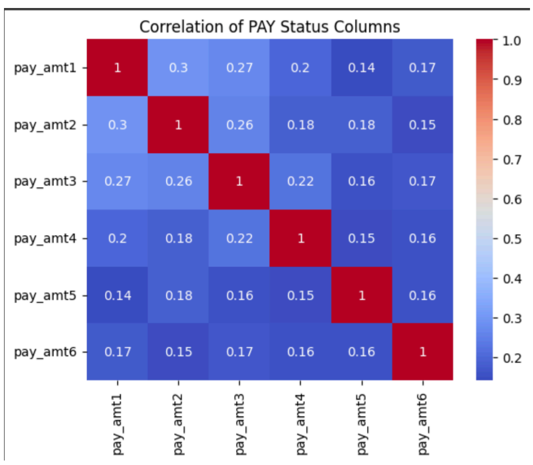
Numerical Variables:

* Histograms were plotted for key numerical features like LIMIT_BAL, age, Avg_Bill_amt, and Pay_To_Bill_ratio, to understand their distributions. This helped identify skewness, outliers, and the spread of values, guiding scaling and modelling decisions.

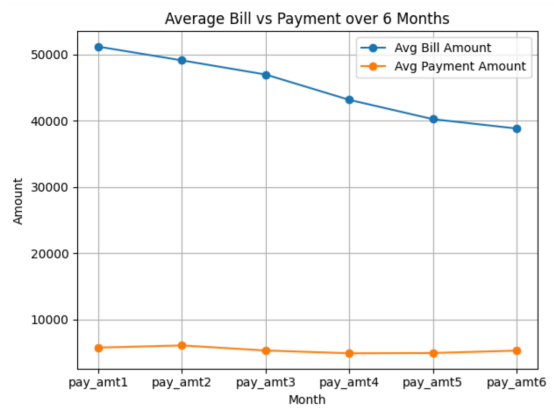


Feature Correlations:

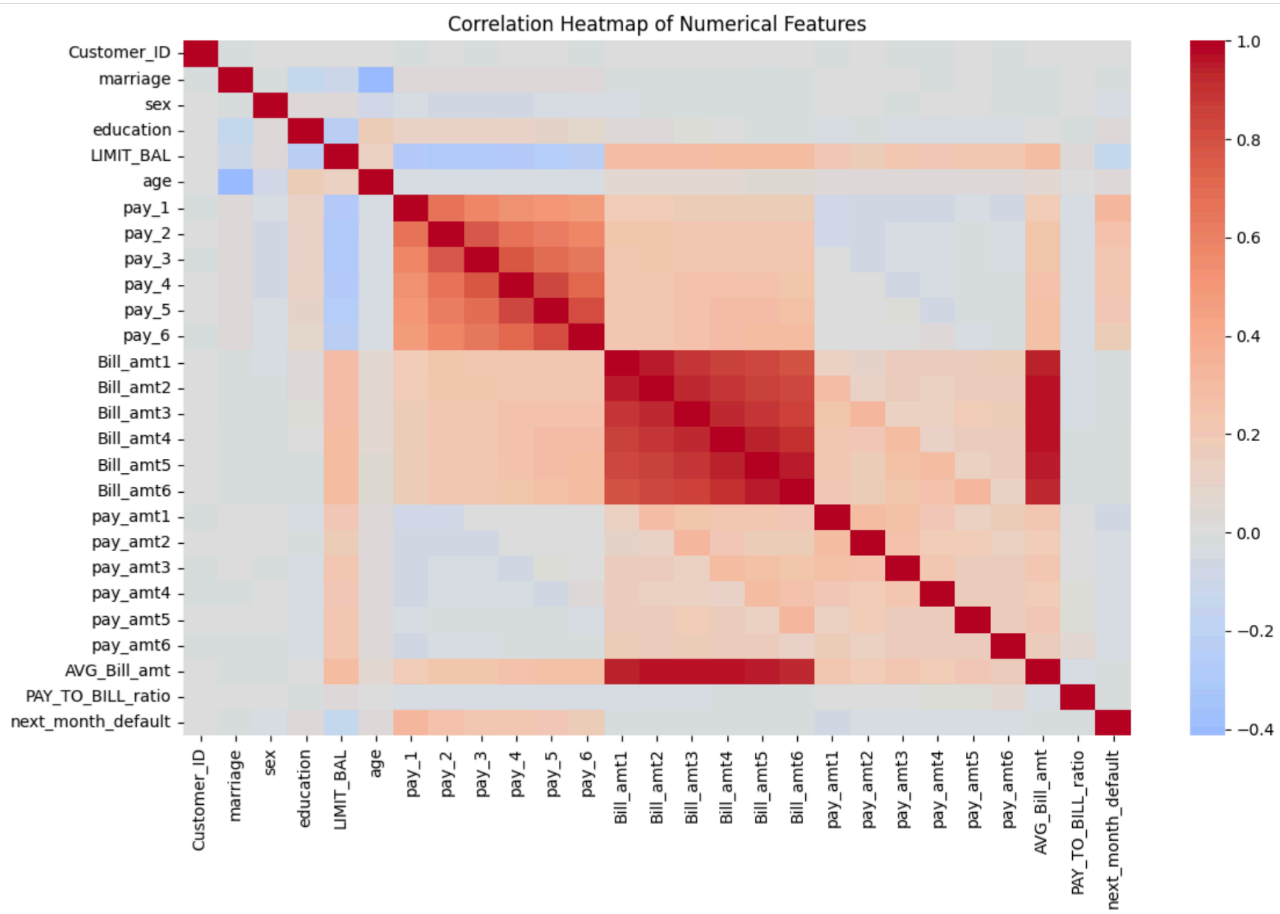
* Bill and payment amounts across months were moderately correlated.
* Delayed payments ('pay_1' to 'pay_6') showed trends associated with default.



A line plot was created to compare the average bill amount and average payment amount over six months. This visualisation helped reveal if users consistently paid less than their bills, highlighting potential risk of default due to underpayment trends.



A correlation heat-map of numerical features was plotted to identify relationships between variables. This helped detect multicollinearity and understand which features are positively or negatively associated with default behaviour.



Custom Features Created:

* 'credit_utilisation_ratio': Avg. bill / Credit limit
* 'count_delayed': Number of delayed months (pay > 0)
* 'repayment_ratio': Total paid / Total billed
* 'avg_utilisation': Mean utilisation over six months

These features were later found influential in improving model performance.

4. Modelling Strategy

To ensure fairness and robustness:

* Train-Test Split: 80-20
* Resampling: Used SMOTE to handle class imbalance.
* Feature Engineering: Incorporated into 'Pipeline' using 'FunctionTransformer'.
* Preprocessing:Numerical features scaled; categorical variables one-hot encoded.

Models Trained:

1. Logistic Regression
2. XGBoost Classifier
3. Decision Tree Classifier
4. Artificial Neural Network (ANN)

Pipelines created accordingly as Logistic Regression and ANN require Scaling and One Hot Encoding while Decision Tree Classifier and XGBoost require Label Encoding

5. **Model Performance (on Test Set)**

Model	Accuracy	Precision	Recall	F2-Score
Logistic Regression.	0.7766	0.4335	0.5906.	0.5507.
XGBoost	0.7147	0.3581	0.6419	0.5540
Decision Tree	0.7220	0.3290	0.4524	0.4208
Neural Network	0.7218	0.3708	0.6764	0.5807

Final ANN model: 3 hidden layers with ReLU, BatchNorm, Dropout. Optimized using Adam and early stopping.

6. **Evaluation Methodology**

Why F2 Score?
F2-score gives more importance to recall, which is crucial in this business case. Missing a defaulter is more costly than flagging a non-defaulter.

Threshold Tuning:
* We used a loop to try thresholds from 0.1 to 0.9.
* Selected the one maximising the F2 score.
* This further improved performance and aligned with business priorities.

7. **Financial Insights**

Key Drivers of Default:

* Delayed Payments: Customers with more months of delay had higher risk.
* Credit Utilization: Higher usage of credit limit was associated with default.
* Low Repayment Ratio: Those who consistently paid less than billed amount were riskier.

8. **Business Implications**

* High Recall Priority: Capturing as many defaulters as possible helps mitigate financial risk.
* Cutoff Flexibility: Threshold can be adjusted based on institution's risk appetite.
* Targeted Interventions: Customers with high credit utilisation or poor repayment ratio can be proactively contacted.

9. **Conclusion and Key Learnings**

This project strengthened our understanding of:

* Data preprocessing and real-world credit datasets
* Feature engineering tailored to financial behaviour
* Importance of resampling and metric selection in imbalanced classification
* Power of threshold tuning for business-specific optimisation

The final ANN model, supported with custom features and threshold tuning, achieved an F2-score of 58.07%, outperforming traditional models.

This approach not only builds an accurate model but also delivers actionable insights for credit risk managers.