CREDIT CARD DEFAULT prediction: A Data-Driven Approach

PRESENTED BY JOVITA ANDREWS

AGENDA

Objective

The goal of this analysis is to predict the likelihood of credit card default by clients on their demographic and financial features. Additionally, we assess model fairness and interpretability to ensure transparency and ethical outcomes

Metholodology

Data Preprocessing:

- Handled missing values.
- Scaled features (eg: credit utilization)

Feature Engineering:Added derived features like
CREDIT_UTILIZATION, PAYMENT_TO_BILL_RATIO

Balancing Dataset:Oversampled minority class to handle
imbalance.

Model Training: Logistic Regression, Random Forest, XGBoost,
Neural Networks

Why these Models?

Logistic Regression: Simple, interpretable baseline
Random Forest: Handles complex patterns, reduces overfitting.
XGBoost: High performance with large datasets
Neural Networks: Superior accuracy for non - linear relationships

models

XGBOOST

- ACCURACY: 77%
- PRECISION: 78%
- RECALL: 77%
- F1- SCORE: 77%

RANDOM FOREST

- ACCURACY: 81%
- PRECISION: 72%
- RECALL: 67%
- F1- SCORE: 80%

MODEL2

model 4

MODEL 1

LOGISTIC REGRESSION

- ACCURACY: 70%
- PRECISION: 76%
- RECALL: 70%
- F1- SCORE: 72%

MODEL 3
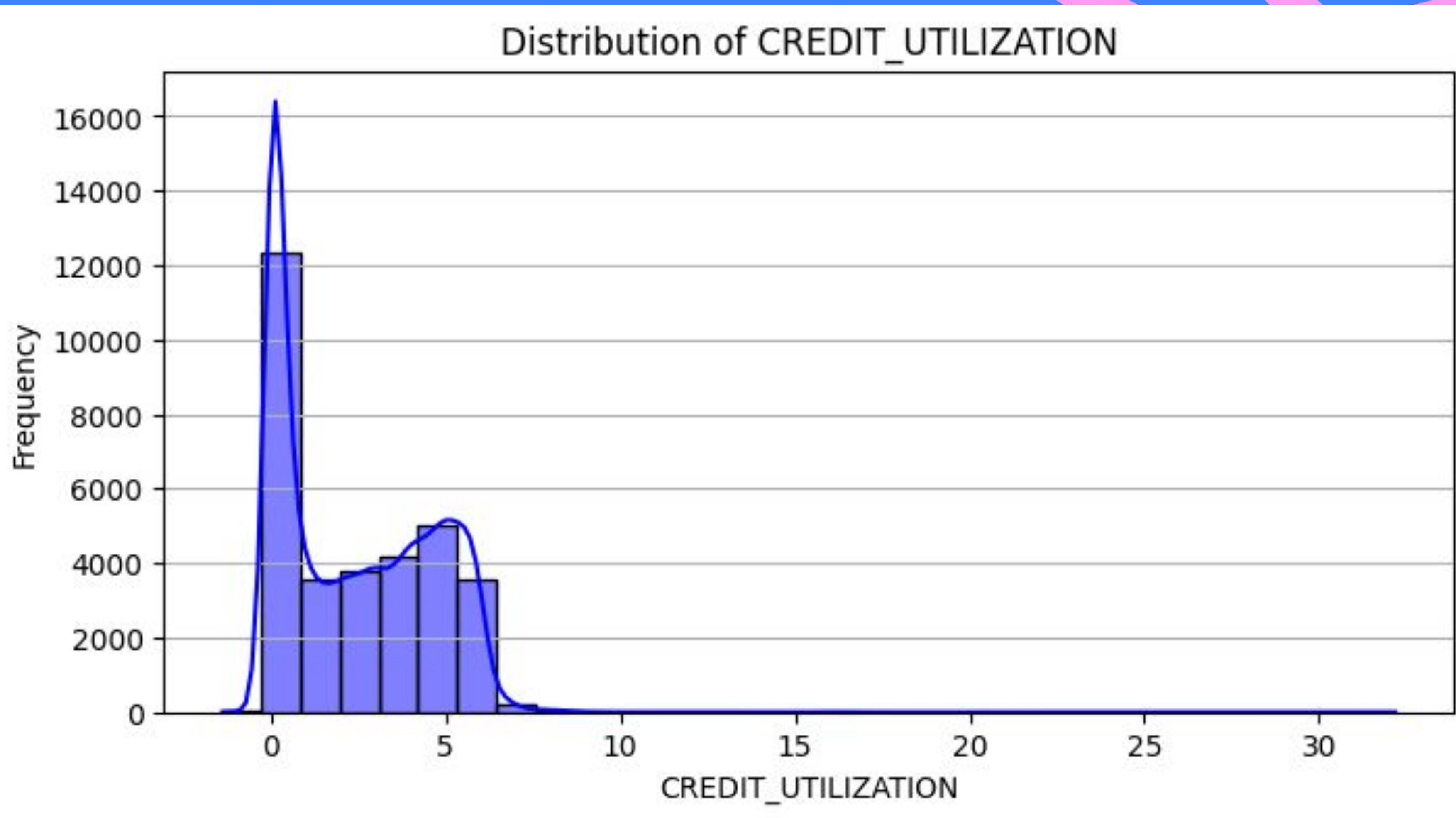
NEURAL NETWORK

- ACCURACY: 77%

Bias ANALYSIS

Demographic Factors: Disparities in predictions based on gender, education and age

Models may inherit societal biases from historical data

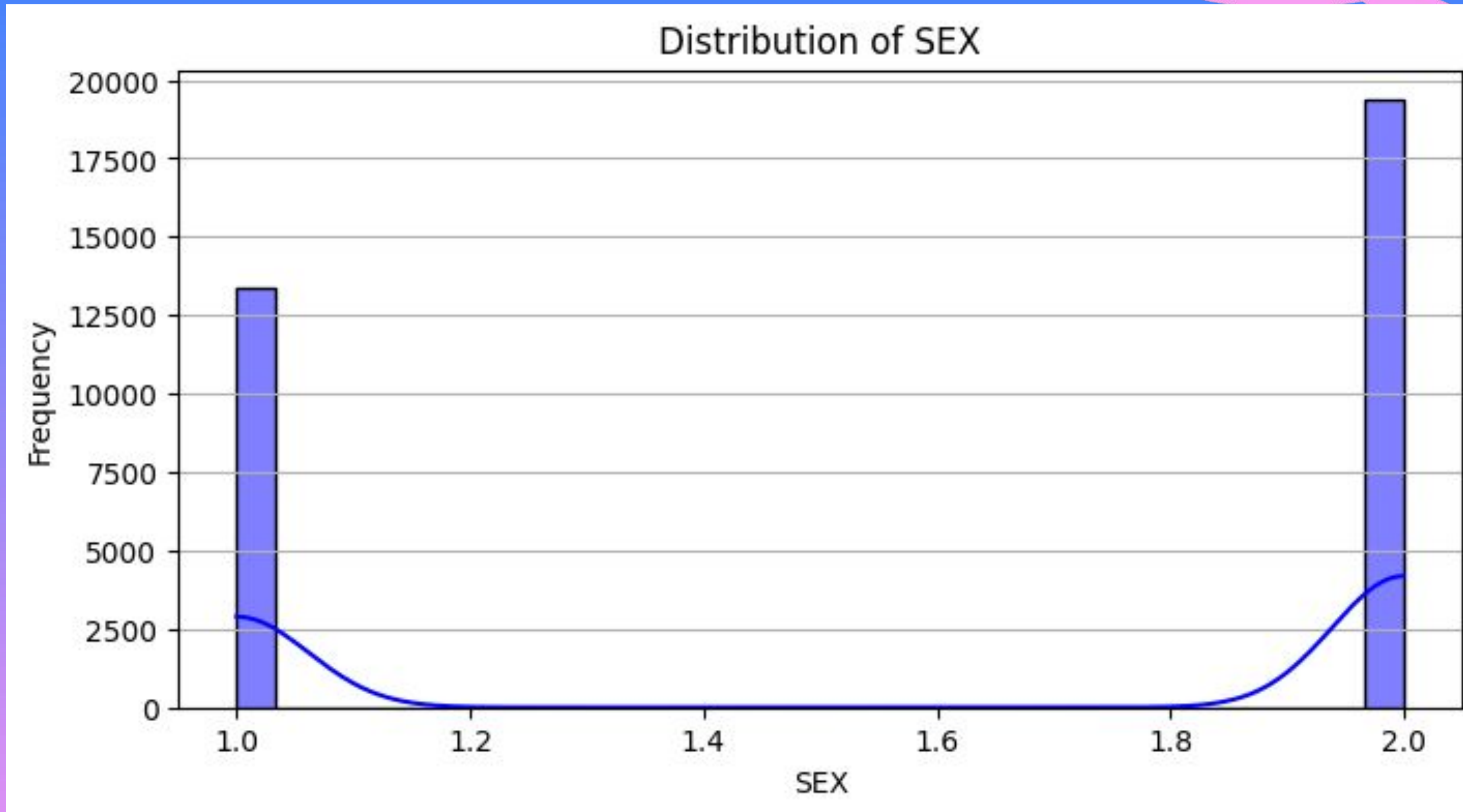Example: Women predicted more likely to default due to systemic biases in credit systems
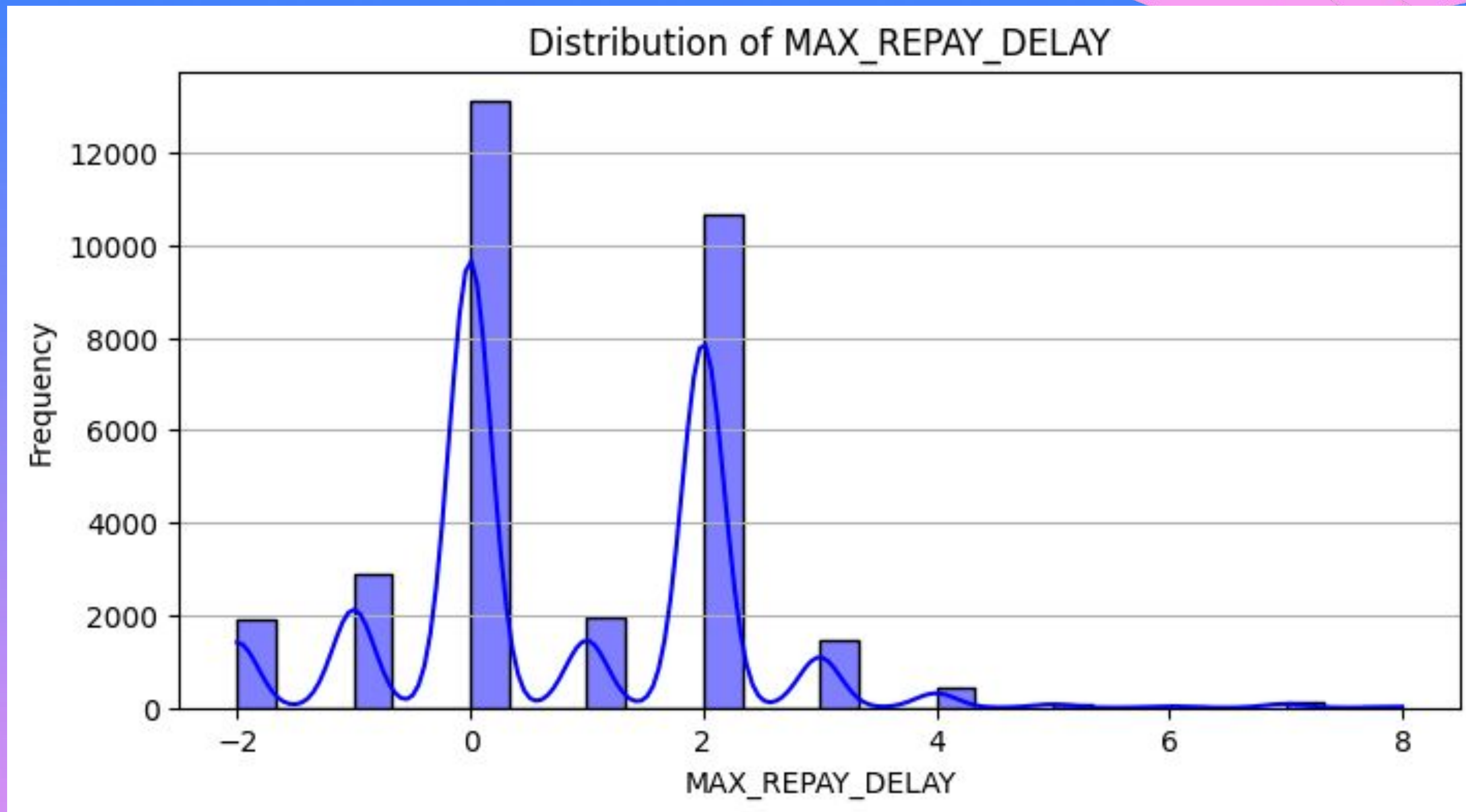
Distribution of CREDIT UTILIZATION

Distribution of sex

Distribution of max REPAY DELAY
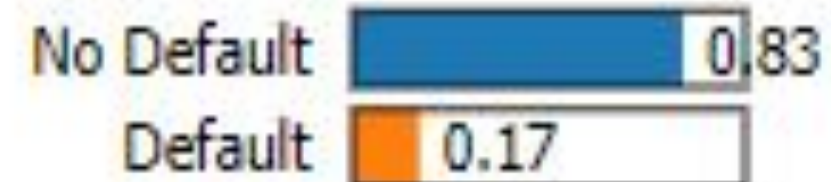

Distribution of MAX_REPAY_DELAY

LIME EXPLANATION FOR SAMPLE INSTANCE

LIME Explanation for Sample Instance:

Prediction probabilities

No Default ████████ 0.83
Default ██ 0.17

No Default | Default

1.00 < EDUCATION ...
  0.06
BILL_AMT1 > 672...
  0.05
1.00 < MARRIAGE <=...
  0.04
-1.00 < PAY_0 <= 0.00
  0.03
BILL_AMT3 > 602...
  0.02
17162.00 < BILL_A...
  0.02
-1.00 < PAY_5 <= 0.00
  0.01
-1.00 < PAY_2 <= 0.00
  0.01
2081.00 < PAY_AMT1...
  0.01
-1.00 < PAY_3 <= 0.00
  0.01
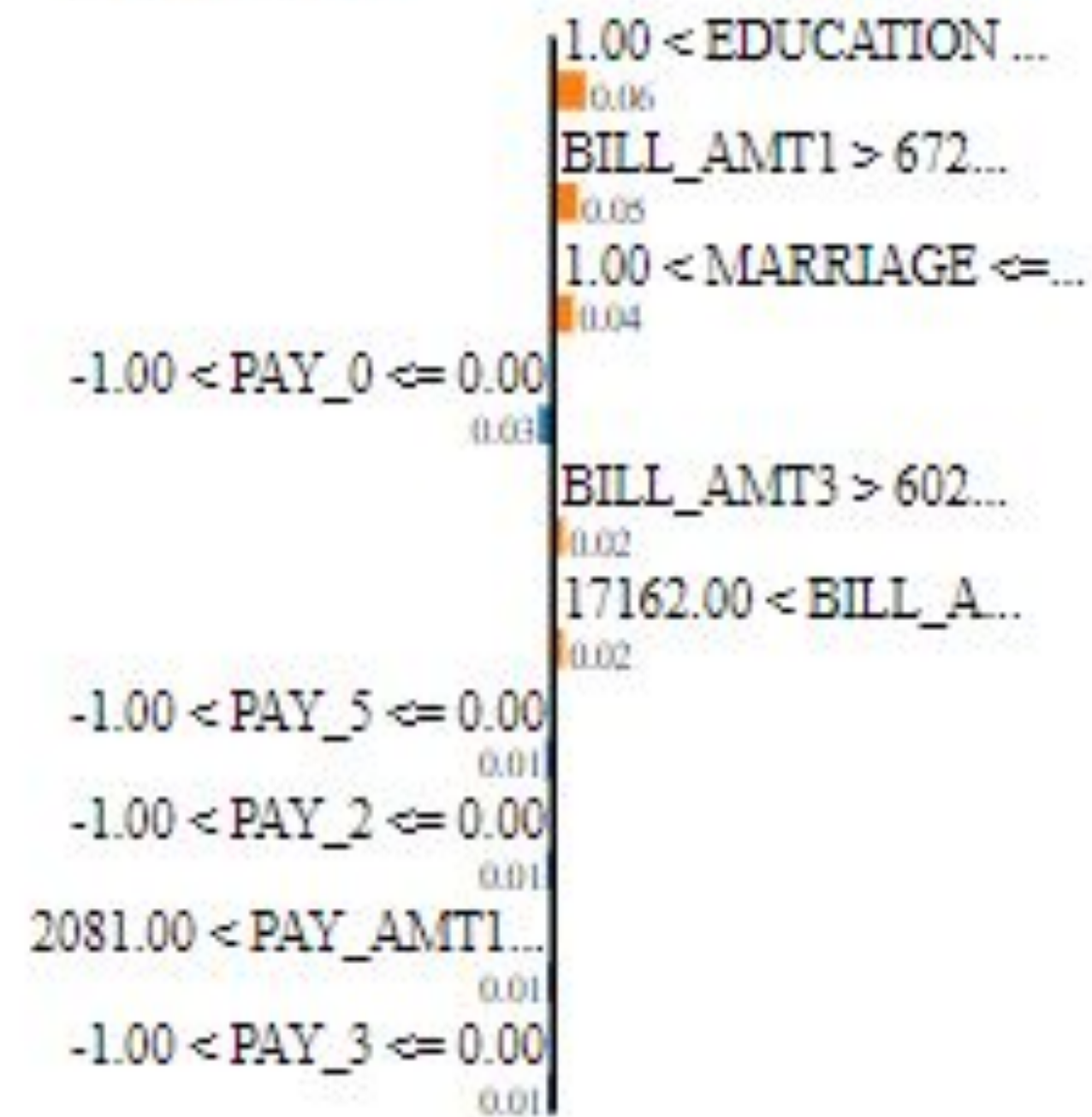
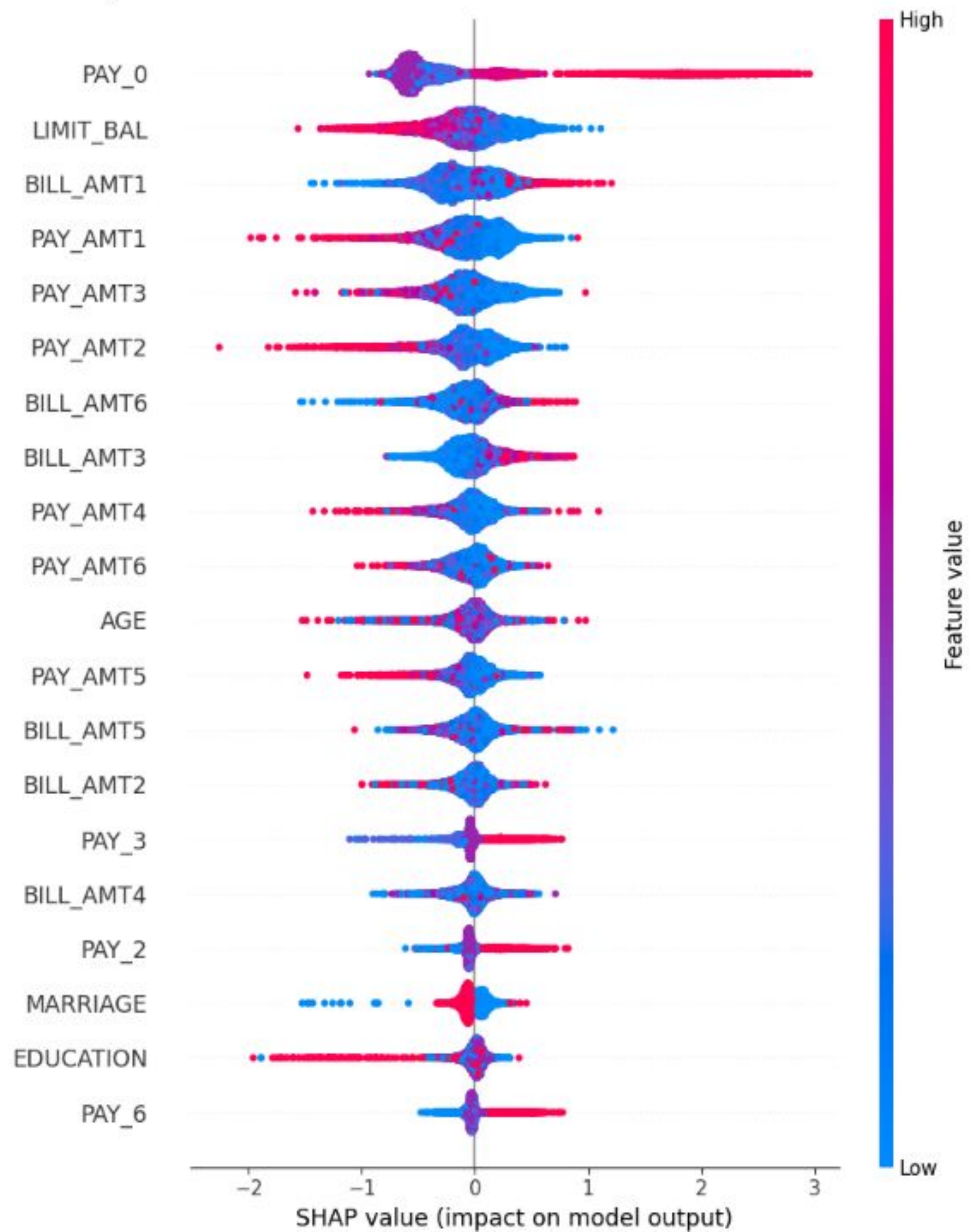| Feature	Value
EDUCATION	2.00
BILL_AMT1	101832.00
MARRIAGE	2.00
PAY_0	0.00
BILL_AMT3	84297.00
BILL_AMT6	44800.00
PAY_5	0.00
PAY_2	0.00
PAY_AMT1	5000.00
PAY_3	0.00

Fairness Metrics:
Demographic Parity Difference: 0.03
Equalized Odds Difference: 0.02

SHAP Summary Plot:

SNAP SUMMARY PLOT

CONCLUSIONS AND RECOMMENDATIONS