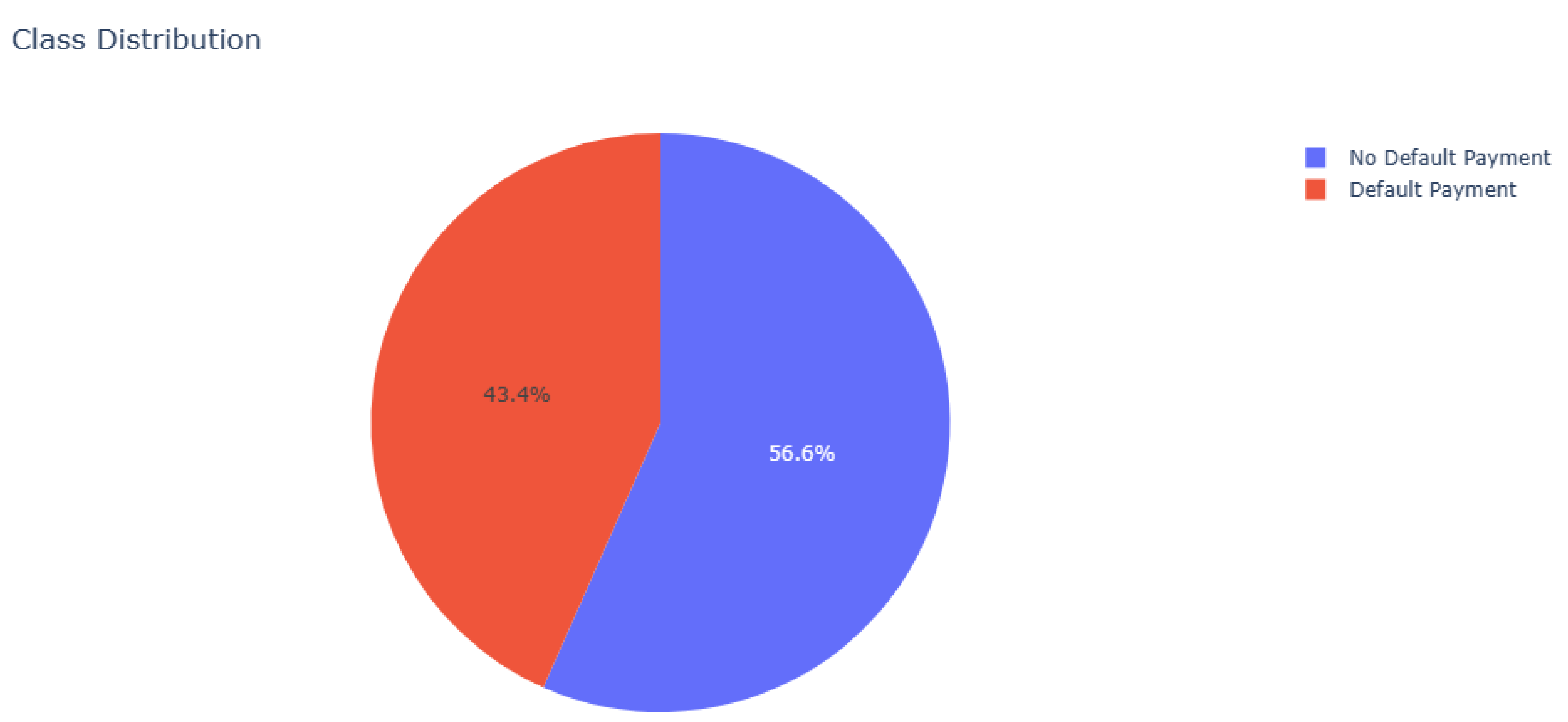
Finance AI-ML Case Study

Credit card is a flexible tool by which a customer can use a bank's money for a short period of time. Predicting accurately which customers are most probable to default represents a significant business opportunity for all banks. Bank cards are the most common credit card type in Taiwan, which emphasizes the impact of risk prediction on both the consumers and banks. This would inform the bank's decisions on criteria to approve a credit card application and also decide upon what credit limit to provide. This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

Using the information given, predict the probability of a customer defaulting in the next month.

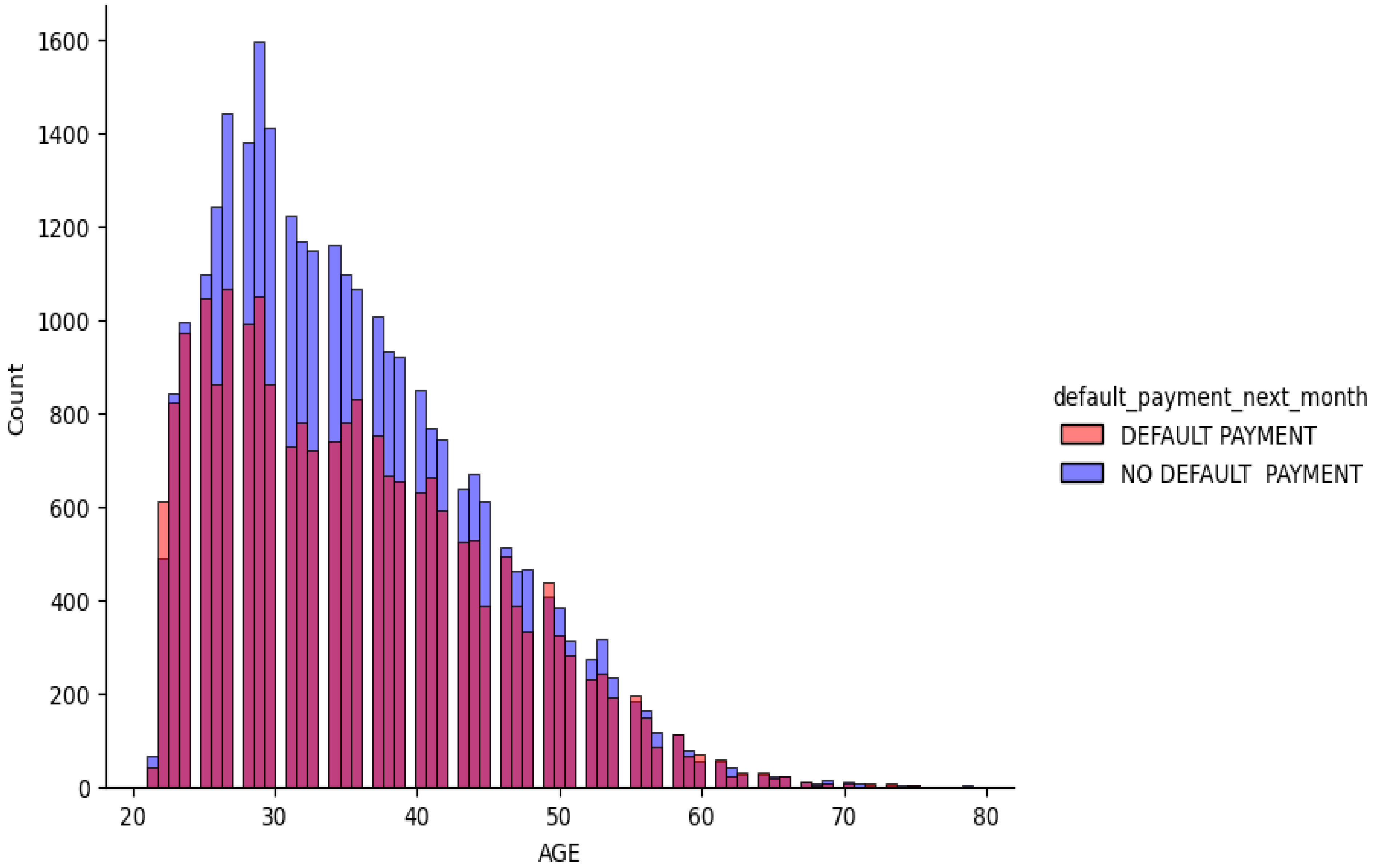
Default Payment = Customer defaulted on payment

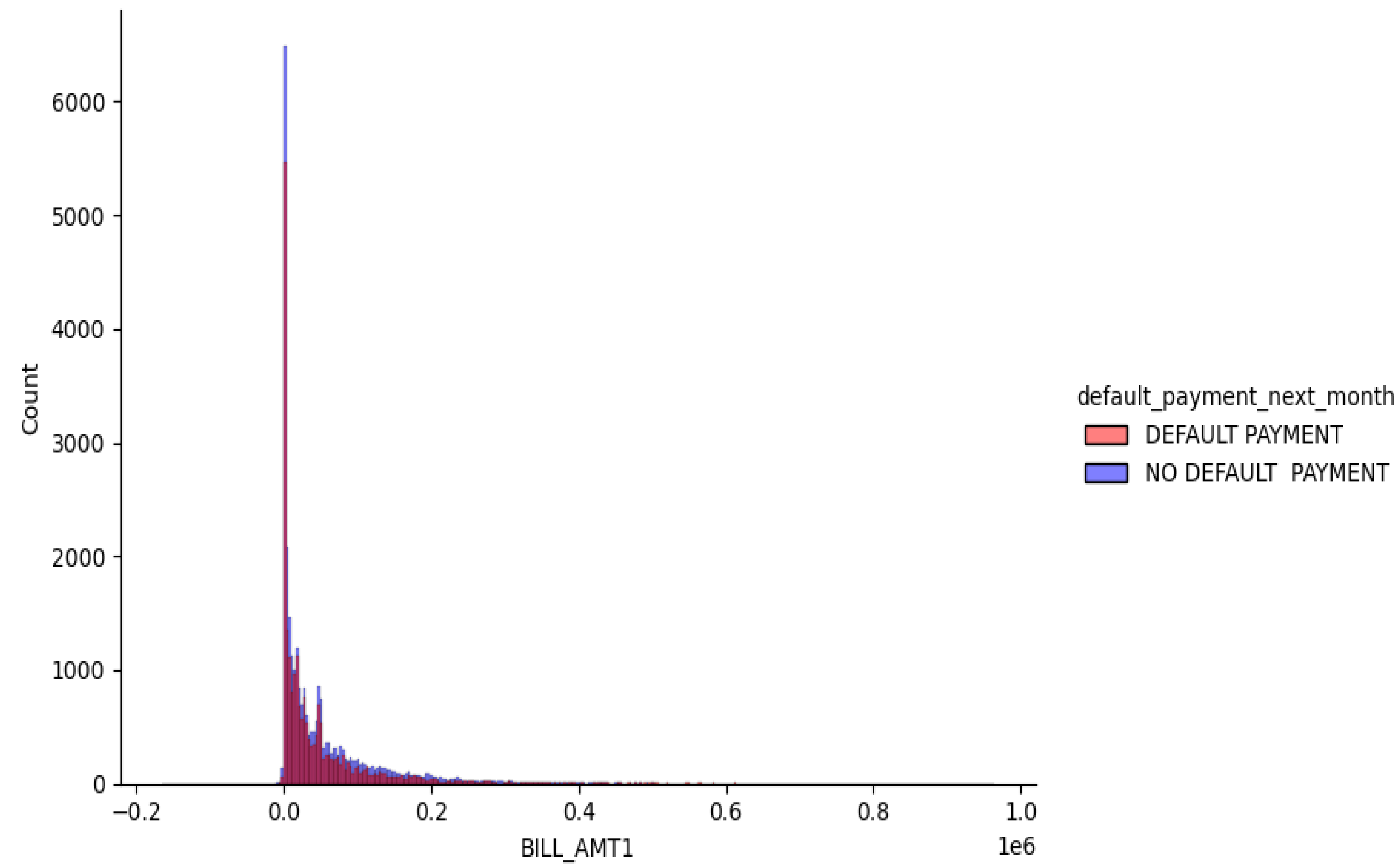
No Default Payment = Customer didn't default on payment



Status	No. of Customers
No Default Payment	28881
Default Payment	22119
Default Payment Conversion	43.4 %
No Default Payment Conversion	56.6 %

Features Responsible : Auto-ML



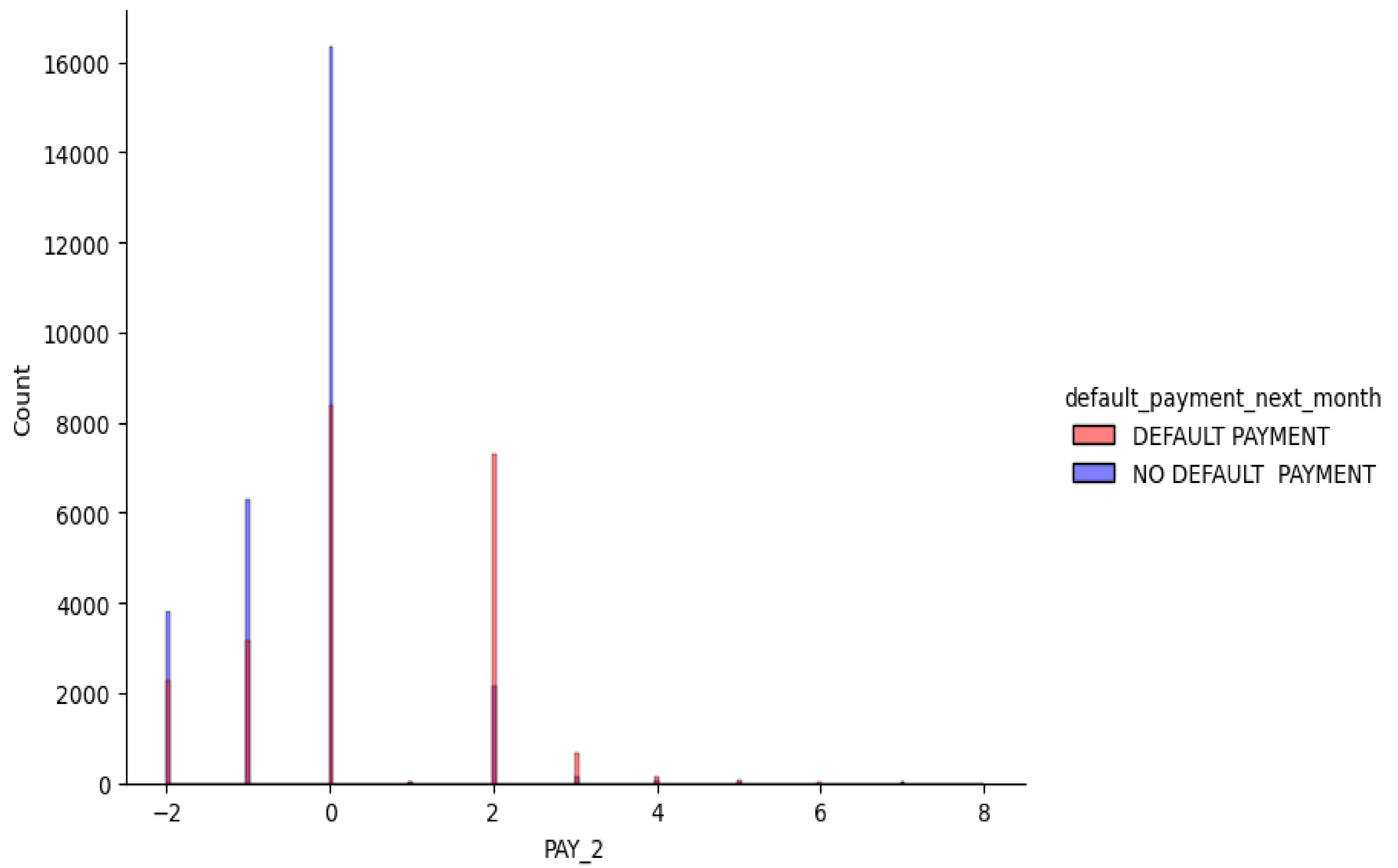


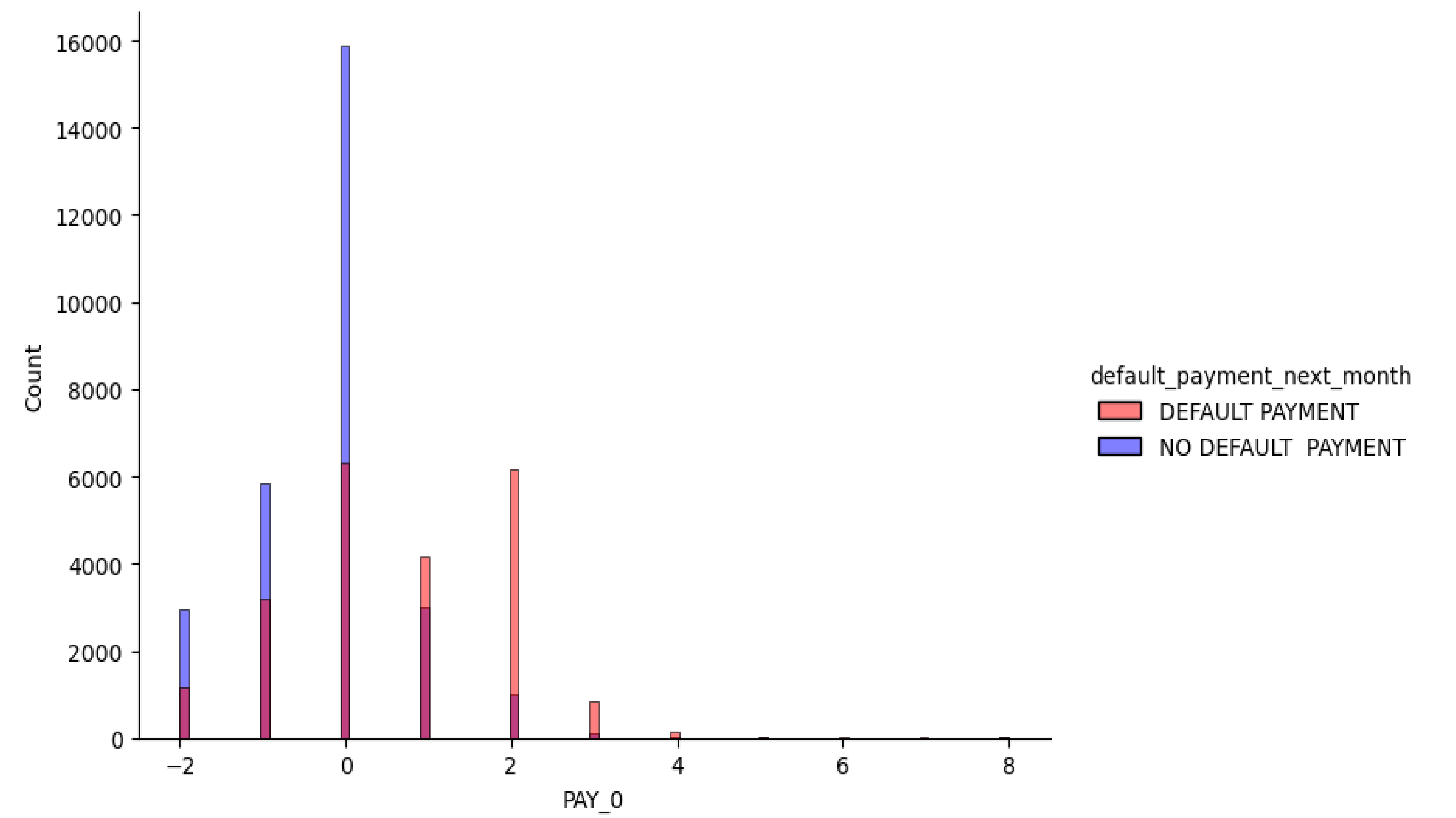
AGE: Age of the client

- 30 to 40 age range shows most not default payments
- Younger customers more prone to taking on debt, such as credit card debt, student loans, which can increase defaulting on payments.
- Older customers have more established financial stability and a longer credit history, which can make them more reliable borrowers.

BILL_AMT1: Amount of bill statement in September, 2005 (NT dollar)

- Generally, customers who have higher bills may be more likely to default on their payments compared to those with lower bills.
- This is because higher bills may indicate that the customer has more debt, and may have a harder time managing their finances.





PAY_2: Repayment status in August, 2005

- If a customer has a history of consistently making their payments on time, they are less likely to default on their payments compared to a customer who has a history of late or missed payments.

PAY_0: Repayment status in September, 2005

- A good repayment history can serve as an indicator of a customer's ability to manage their finances and meet their financial obligations, which can reduce their risk of defaulting on their payments.

Auto-ML Methodology Results

Algorithms	Test Accuracy (25 Percentile)	Test Accuracy (50 Percentile)	Test Accuracy (75 Percentile)	Test Accuracy (90 Percentile)
Gaussian Naive Bayes	49.62	49.24	50.10	49.56
Logistic Regression	58.63	58.68	58.02	58.35
Random Forest	92.48	92.78	93.79	93.59
XGBoost	80.17	80.47	81.23	80.98
MLP	56.55	56.75	56.01	56.36
No. of Features	8	16	24	29
Avg. Accuracy	67.49	67.58	67.83	67.77

- Based on our observations, 75 percentile has the best average accuracy of 67.83.
- Random Forest is the best performing algorithm with 93.79 % accuracy at 75^{th} percentile

Auto-ML Conclusion

- In conclusion, the application of AI-ML in the finance industry has the potential to greatly improve various processes and enhance decision-making capabilities. By leveraging AI-ML algorithms, financial institutions can automate tasks and reduce the potential for human error. The dataset contains 51000 records with 11 Categorical and 14 Numerical Features.
- For classification, models were created with algorithms using Auto-ML techniques like: Gaussian Naïve Bayes, Multi-layer Perceptron, Random forest , XGBoost and Logistic Regression. With these models, performance measurement values were obtained for feature sets of 8, 16, 24 and 29.
- The Auto-ML algorithms were able to predict whether the payments was default or not based on their features with an average accuracy between 65% – 70% and helped to identify features that determine the default payment rate and predict whether the customer has a high chance to make a default payment or not. The major factors are Age, Amount of bill statement in September, Repayment status in August and Repayment status in September.
- In summary, data science and ML are transforming the Finance industry by enabling Auto-ML algorithms that can make quick decisions on underwriting and credit scoring and save companies both time and financial resources that are used by humans.

Sensitivity Analysis

Actual Data Values

AGE	BILL_AMT2	PAY_0	PAY_2	PAY_3	DEFAULT_PAYMENT_NEXT_MONTH
22.0 59 %	7948.0 -60 %	1.0 100 %	2.0 50 %	0.0 50 %	1.0

Default Payment Next Month - 1 : Is a Default Payment

Adjusted Data Values

AGE	BILL_AMT2	PAY_0	PAY_2	PAY_3	DEFAULT_PAYMENT_NEXT_MONTH
43.24 ↑ (21.24)	6752.8 ↓ (-1195.2)	-1.0 ↓ (-2.0)	1.0 ↓ (-1.0)	-1.0 ↓ (-1.0)	0.0

Default Payment Next Month - 0 : No Default Payment

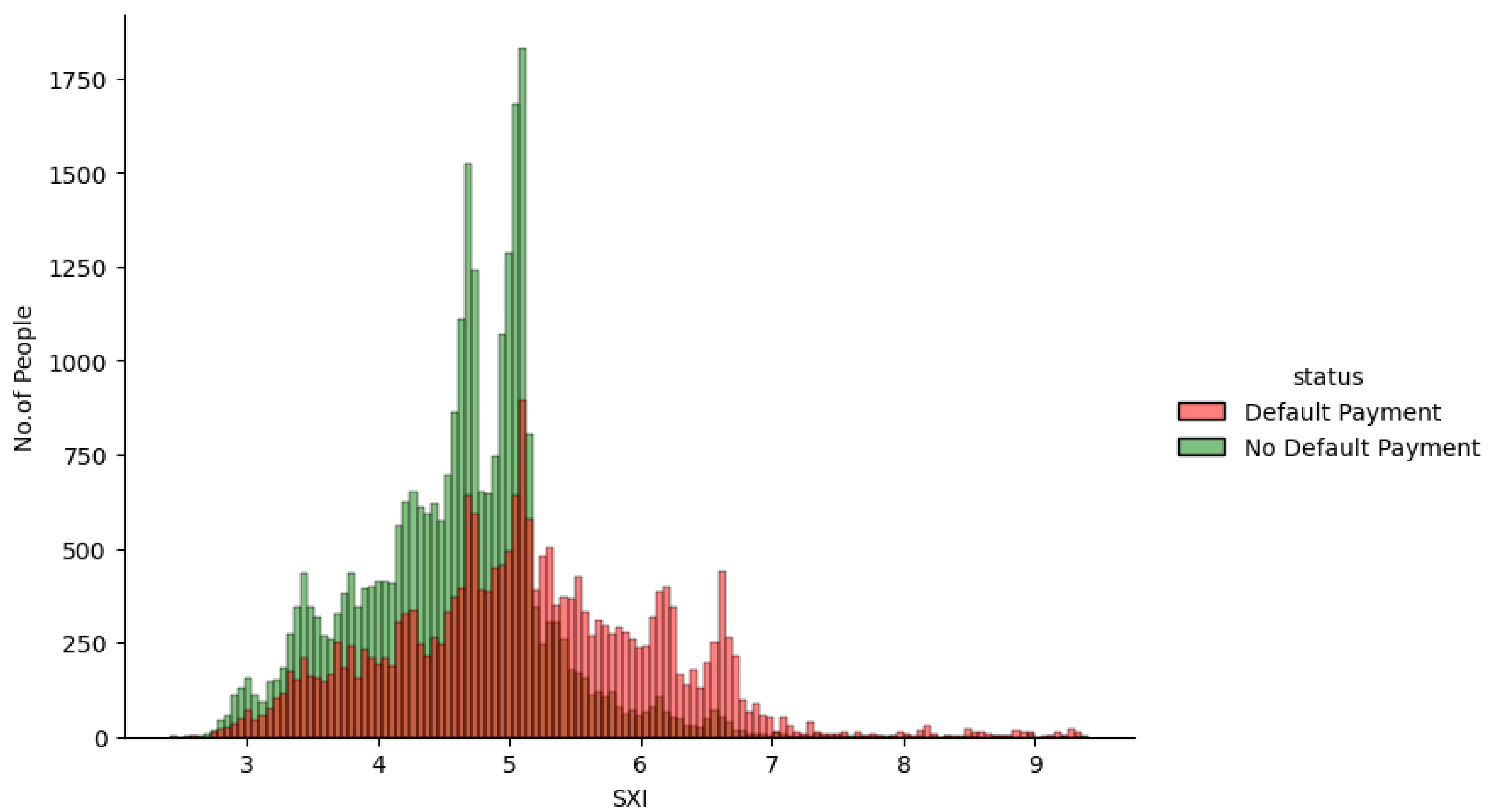
- The top 5 features :
  - AGE : Age of the customer.
  - BILL_AMT2: Amount of bill statement in August (dollars)
  - PAY_0: Repayment status in September
  - PAY_2: Repayment status in August
  - PAY_3: Repayment status in July

For this customer we can see:

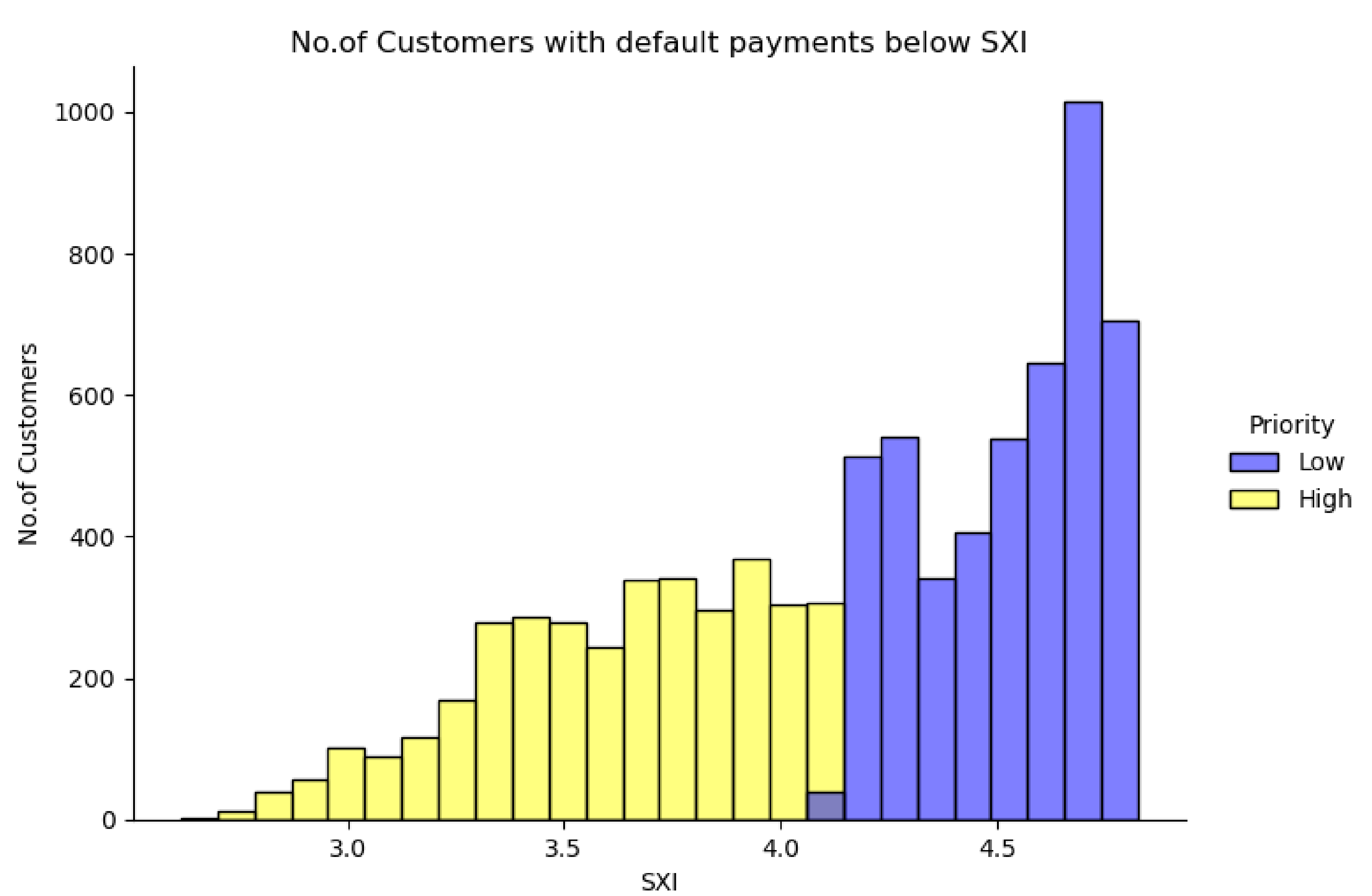
- If the age of the customer with 59% increase from the actual .
- If the amount of bill statement with 60 % decrease from the actual .
- If the repayment status in September with 100 % decrease from the actual.(-1 is paid duly)
- If the repayment status in August with 50 % decrease from the actual.(1 is payment delay for one month)
- If the repayment status in July with 50 % decrease from the actual.(-1 is paid duly)

Which will lead this customer to not to default their payment.

SXI Method



- Brown – Mix of default payments and no default payments
- Pink – Default Paymetns
- Green - No Default Payments



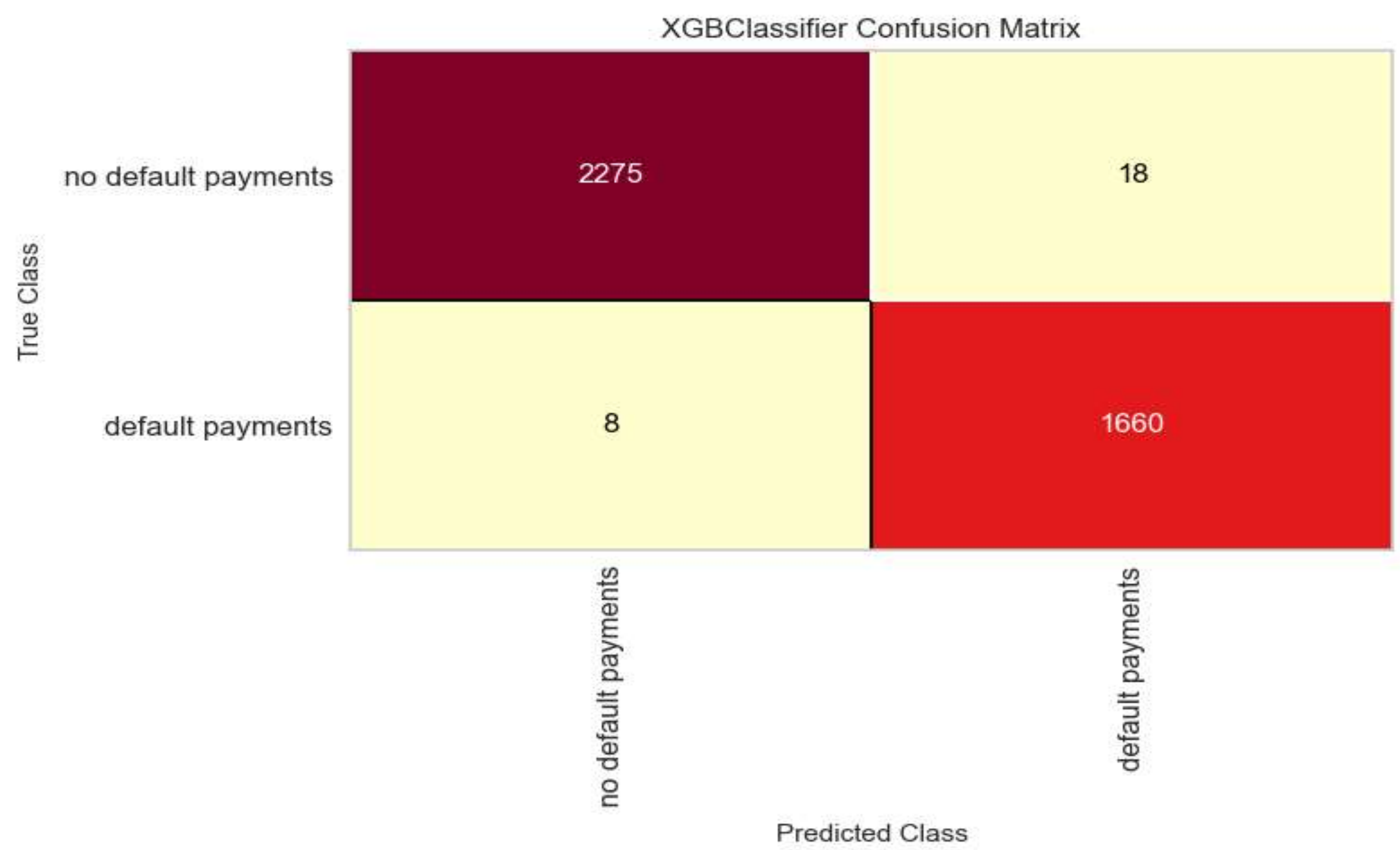
SXI	4.83
Top SXI	9.41
Minimum SXI	2.41
No. of Default payments above SXI	13751
No. of Default payments below SXI	8368
No. of No Default payments above SXI	11435
No. of No Default payments below SXI	17446
SXI Model Accuracy	99.3

As SXI increases, default payment conversion increases.

Default Payments below SXI

- In order to decrease the default payment conversion, these customers are the one to focus on, because they have more chance not to default on their payment
- The average SXI on "Default Payment who are below SXI" is 4.1. Here we further categorize ,so customer's SXI below 4.1 are to be mostly retargeted in future.
- Analyzing the trend and behavior patterns of these customers will improve getting better, loyal and more adequate customers.

SXI Method Accuracy is 99%



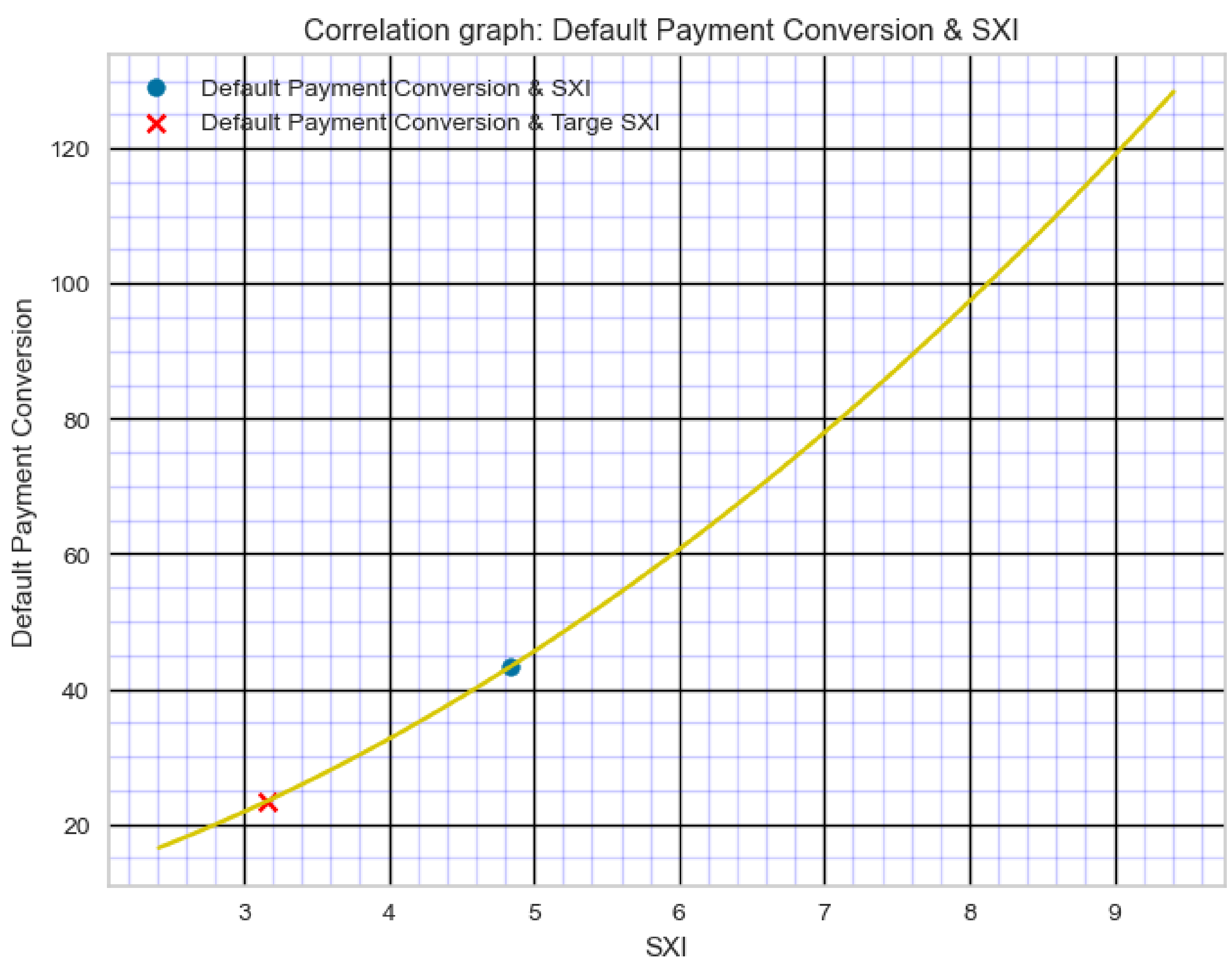
- Actual and Predicted was No Default Payment (TP) : 2275
- Actual and Predicted was Default Payment (TN) : 1660
- Actual No Default Payment and Predicted Default Payment (FN): 18
- Actual Default Payment and Predicted No Default Payment (FP) : 8

Train Records	Test Records	Actual Train count for default payment	Actual train count for no default payment	Actual test count for default payment	Actual test count for no default payment	Predicted test count default payment	Predicted test count no default payment	Precision rate	Recall rate	Model Accuracy
15842	3961	6700	9142	1668	2293	1678	2283	98.9	99.5	**99.3**

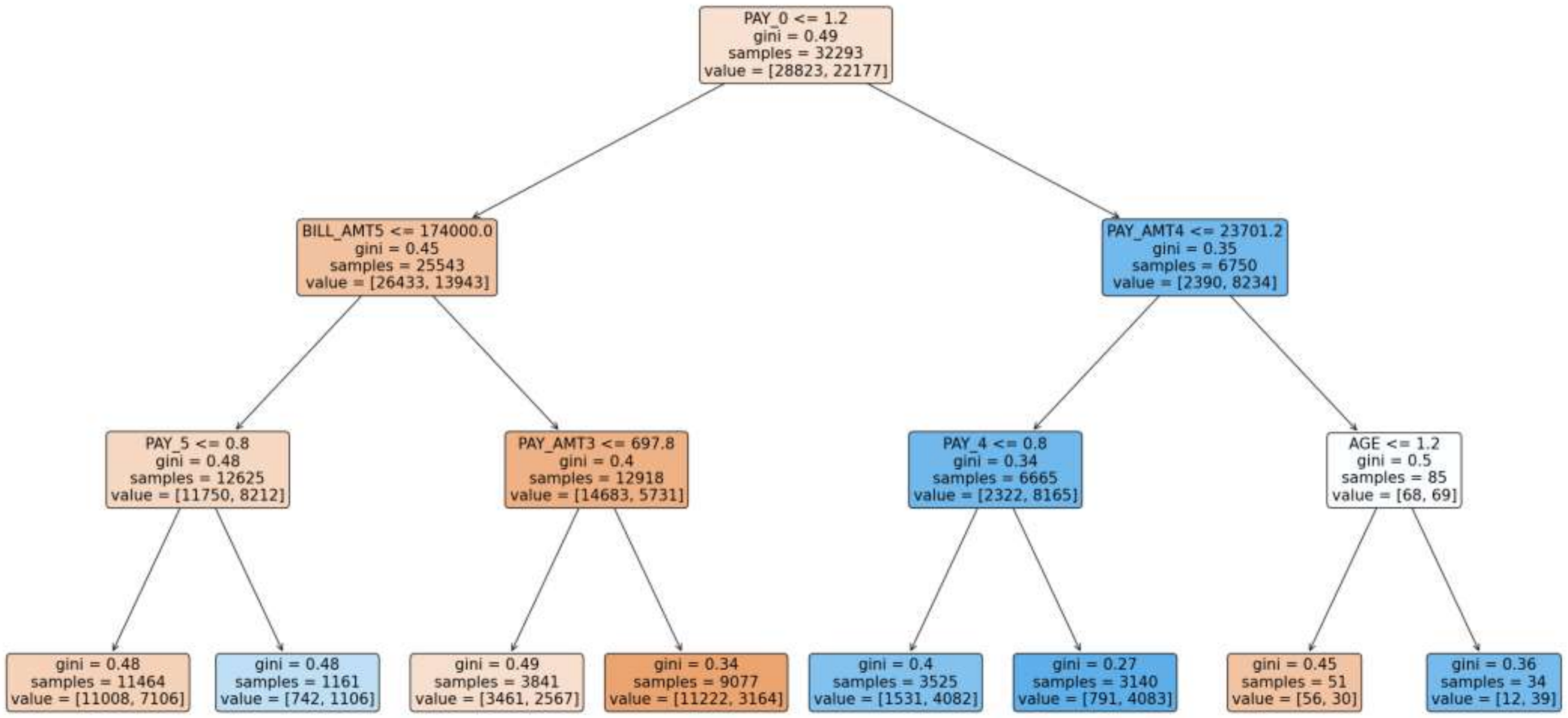
- **Precision rate**: Precision is defined as the ratio of actual No Default Payments (True Positive) to a total number of predicted No Default Payments. ***TP / (TP + FP)***
- **Recall rate**: The recall is calculated as the ratio between the Actual numbers of No Default Payments to the total number of wrongly predicted No Default Payments as Default Payments plus actual number of No Default Payments. ***TP / (TP + FN)***
- **Model Accuracy**: It is the fraction of predictions where the model got right. ***(TP + TN) / (TP + FP + TN + FN)***

SXI Method - Conclusion

SXI	4.83
Target SXI	3.16
Default Payment Conversion	43.4
Target Default Payment Conversion	23.37



The correlation between SXI and Default Payment Conversion is ***0.99***



Tree - Interpretation

- Repayment status in September= pay duly or payment delay for 1 month
- Amount of bill statement in June in dollars < = 174000
- Amount of previous payment in July in dollars >= 697.8
- Lead to a decrease in default payments.