Credit Default Risk Classification

Lifeline Creditors

Project Overview

- Objective: Predict whether a credit card client will default on payment next month.
- Dataset: UCI Credit Card dataset containing 30,000 client records.
- Model: Decision Tree Classifier, with performance metrics including accuracy, AUC, and confusion matrix.

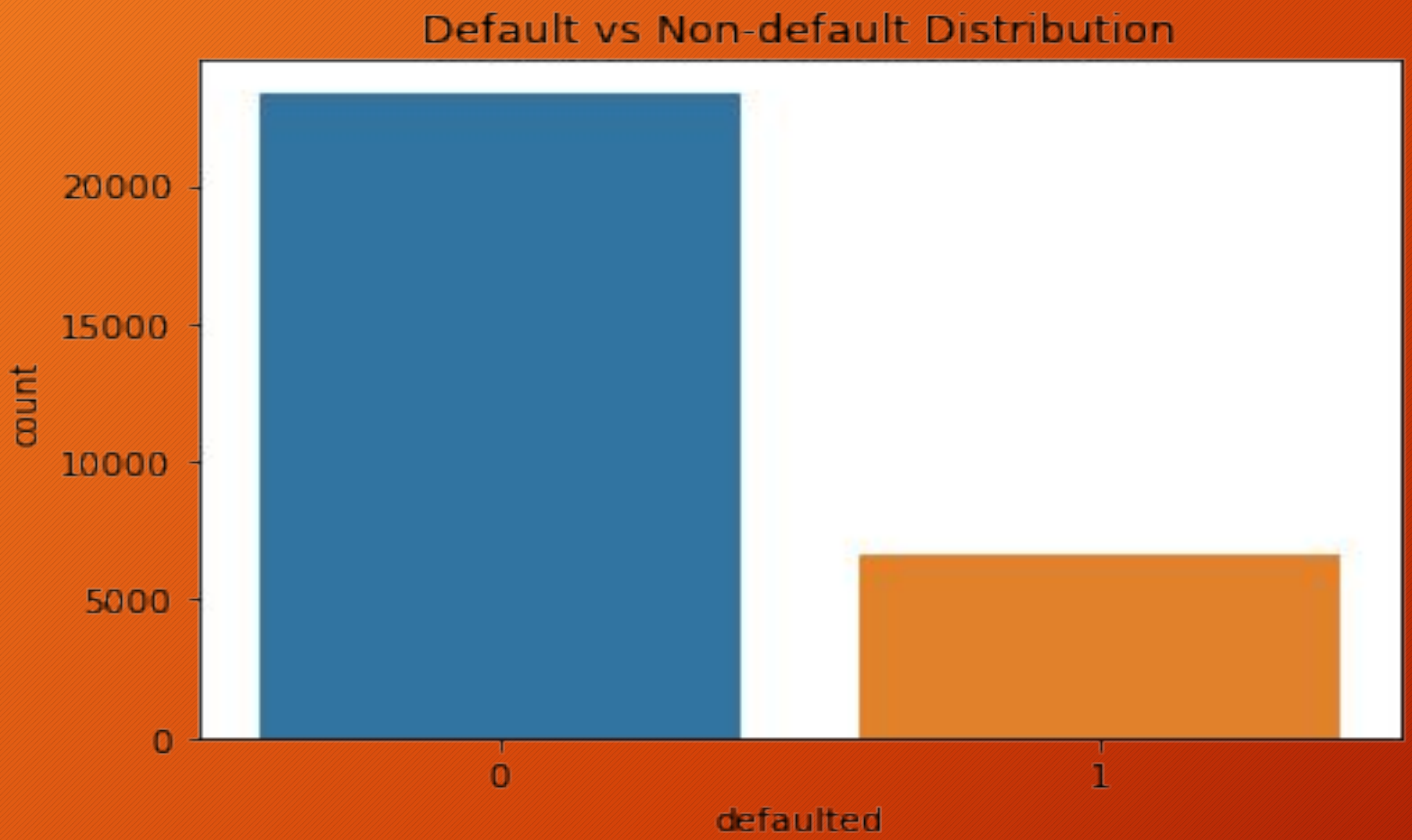
Key Features Considered

- \- Credit Limit
- \- Education Level
- \- Marital Status
- \- History of Past Payment
- \- Monthly Bill & Payment Amounts

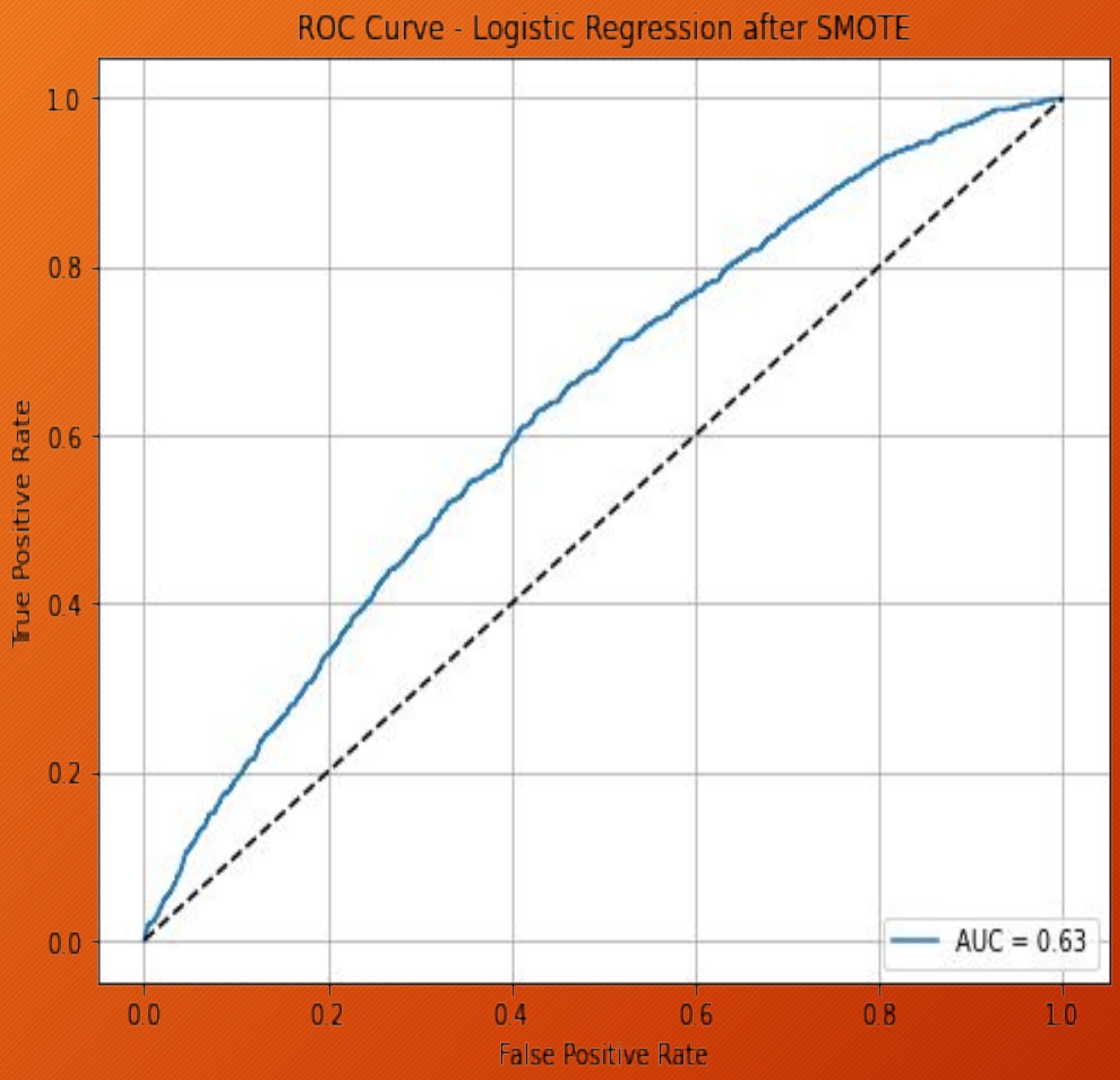
Model Performance

- The decision tree model was evaluated using:
- - Accuracy Score
- - ROC Curve and AUC Score
- - Confusion Matrix

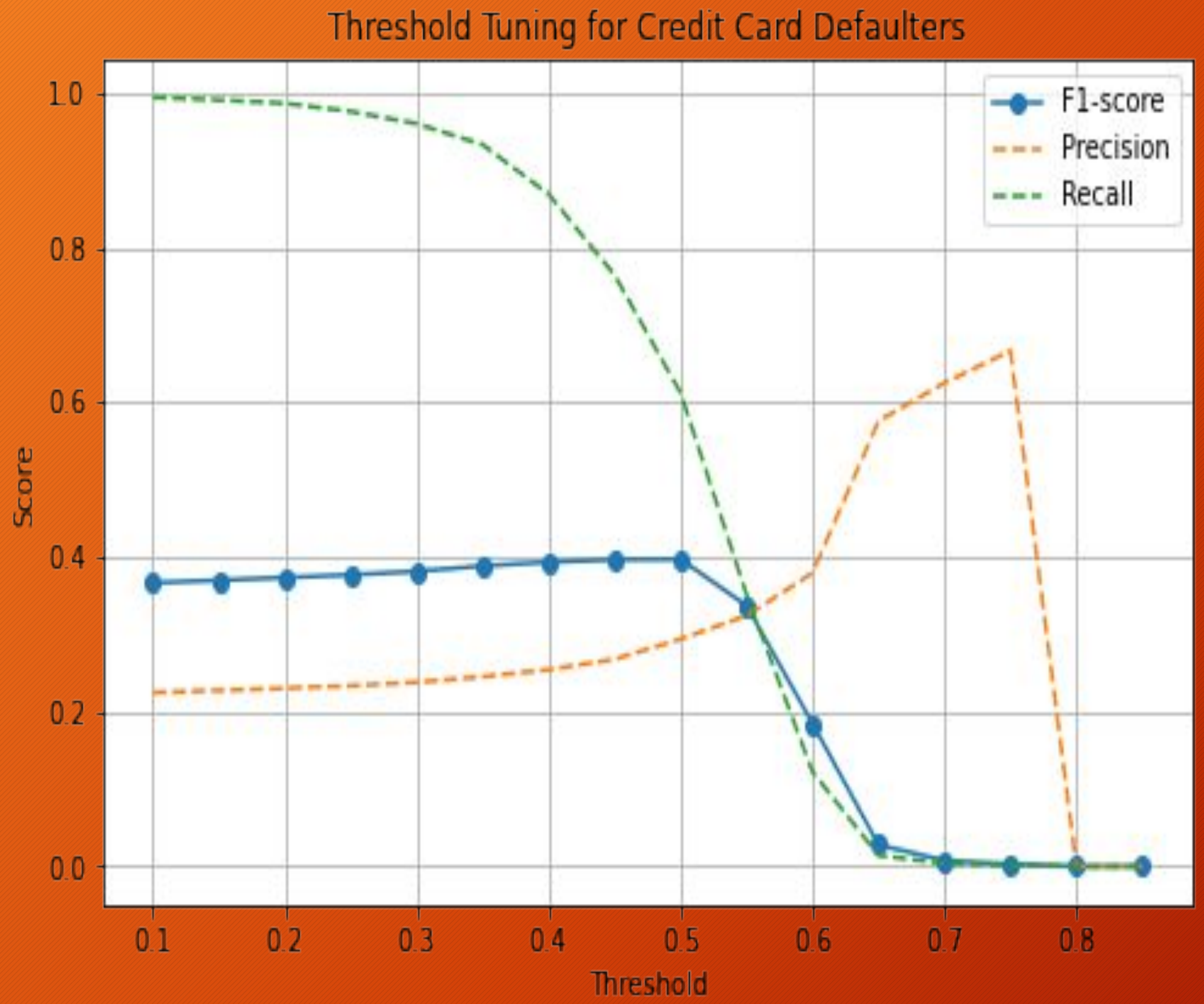
- Threshold tuning was used to improve the balance between precision and recall.



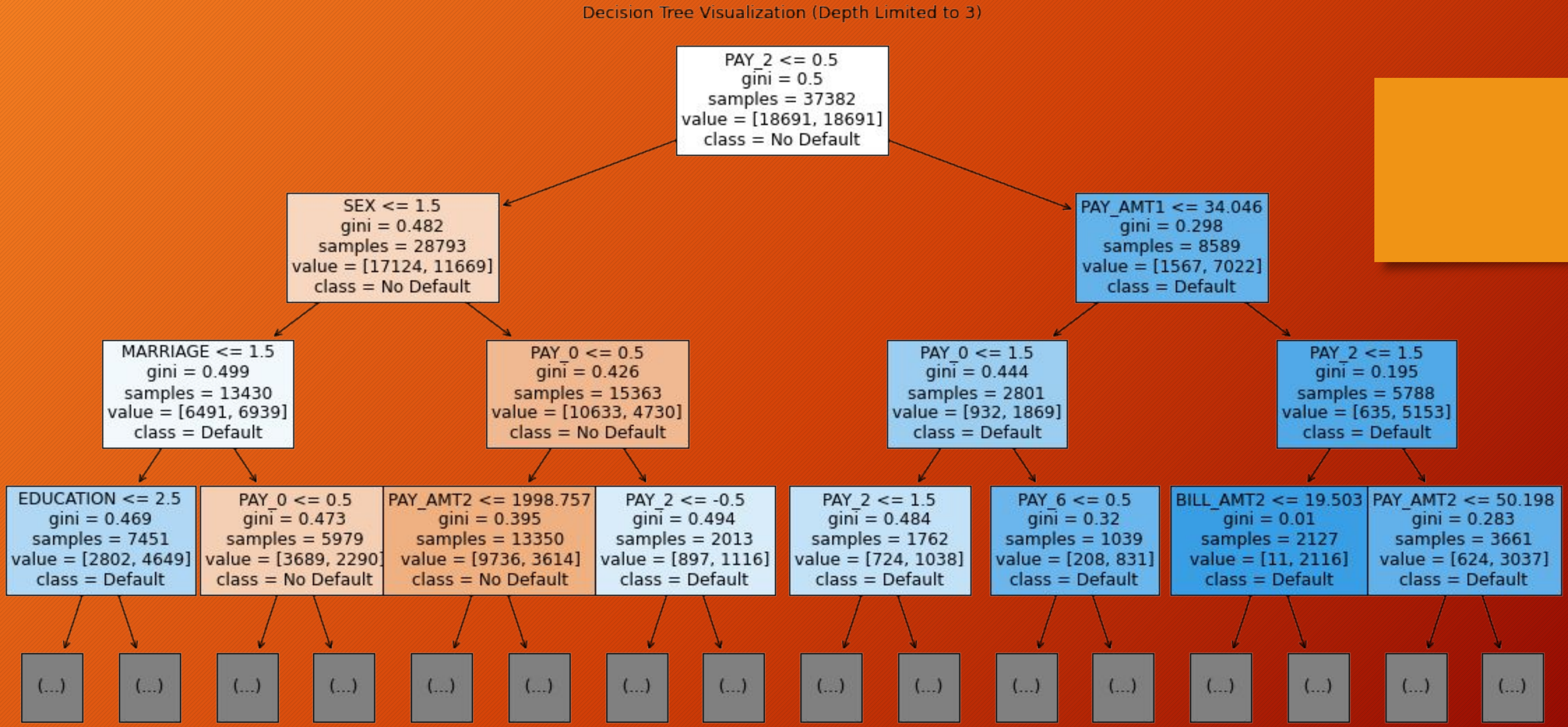
The chart shows the proportion of defaults against the non-defaulted. The classes have a varying contrast; therefore, we introduce synthetic data points to the minority class to reduce the imbalance and make analysis easier.



After fitting a logistic regression, we plot the ROC curve that shows the model's performance. The AUC (area under the curve) is 0.63. This shows that there is some predictive power beyond 0.5 (**the dotted line**) that indicates random guessing. This is good, but it should be better to detect and predict defaulters more aggressively.

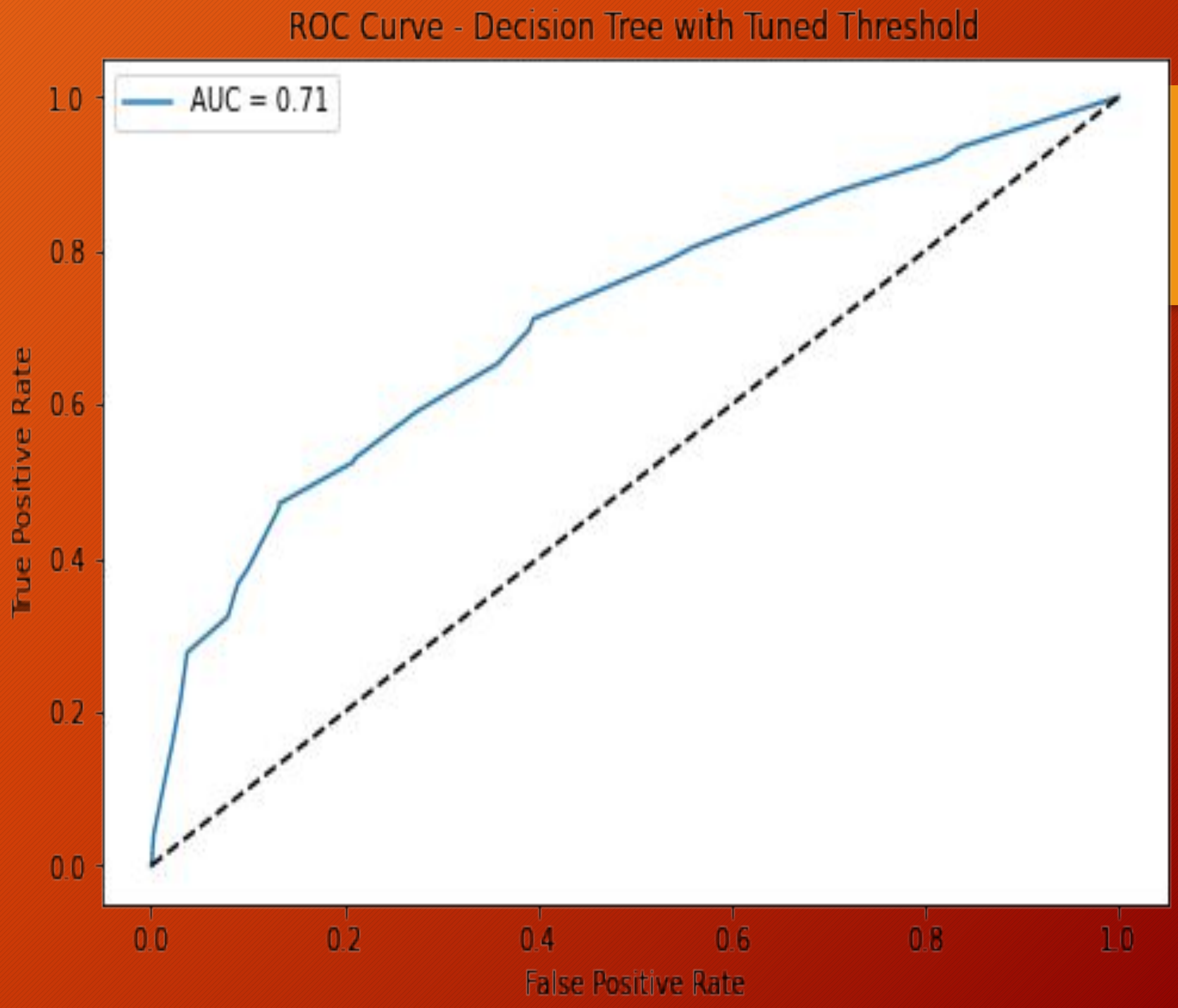


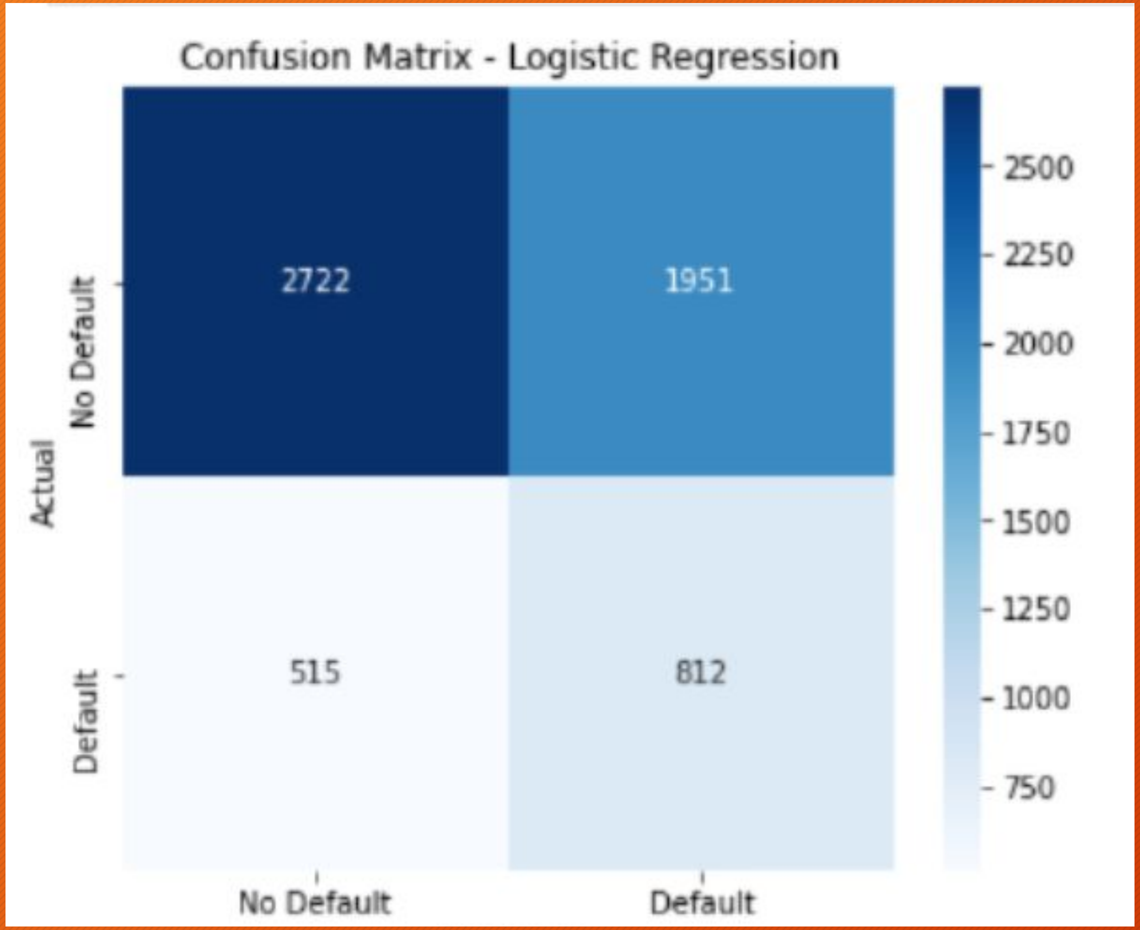
To improve overall model performance, we tune the decision-making threshold to make it more sensitive. The optimal threshold would be where the curves intersect, which is 0.37.

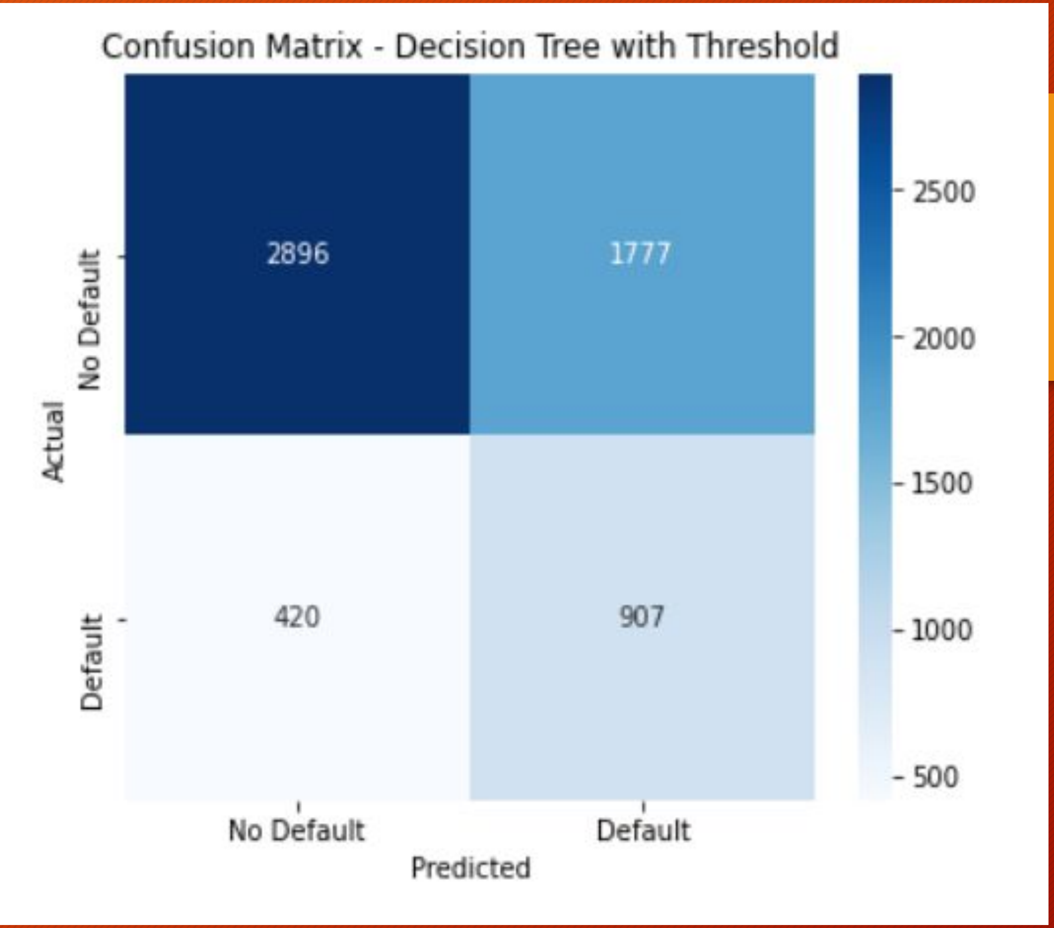


With our new threshold, we train the model with a Decision tree classifier, which highlights features that determine whether a person is likely to default.

The AUC (area under the curve) is 0.71. This shows that there is more predictive power than the previous of 0.63. This is better at detecting and predicting defaulters more aggressively.







After tuning the decision threshold, the confusion matrix changes from the one on the left to the one on the right. The new matrix can be interpreted as:

True Positives (TP) = 907 → Correctly predicted defaults.

True Negatives (TN) = 2896 → Correctly predicted non-defaults.

False Positives (FP) = 1777 → Non-defaults incorrectly predicted as defaults.

False Negatives (FN) = 420 → Defaults missed by the model. Overall, the model has improved its precision and Recall, which aligns with the company's goals

Insights and Recommendations

- -Clients with low credit limits and inconsistent past payments should be flagged for review.
- Younger clients and those with high monthly bill statements may require stricter approval conditions.
- -Clients with lower education levels have a higher loan default rate, suggesting an element if financial literacy in funds management.
- Married clients have a better loan repayment rate compared to single clients. This could be due to the fact of having two contributors in a household easing the burden on financial resources.
- Incorporate this model into your approval system for early risk identification.