Moneza

Automating Loan Application Processing

MIKE KIPTOCH

6th July 2024

Moneza

INTRODUCTION

- Manual processing of loan applications is costly.
- Aim: Develop a machine learning model to automate approval/rejection.
- Goal: Minimize labor costs and financial losses from incorrect decisions.

Moneza

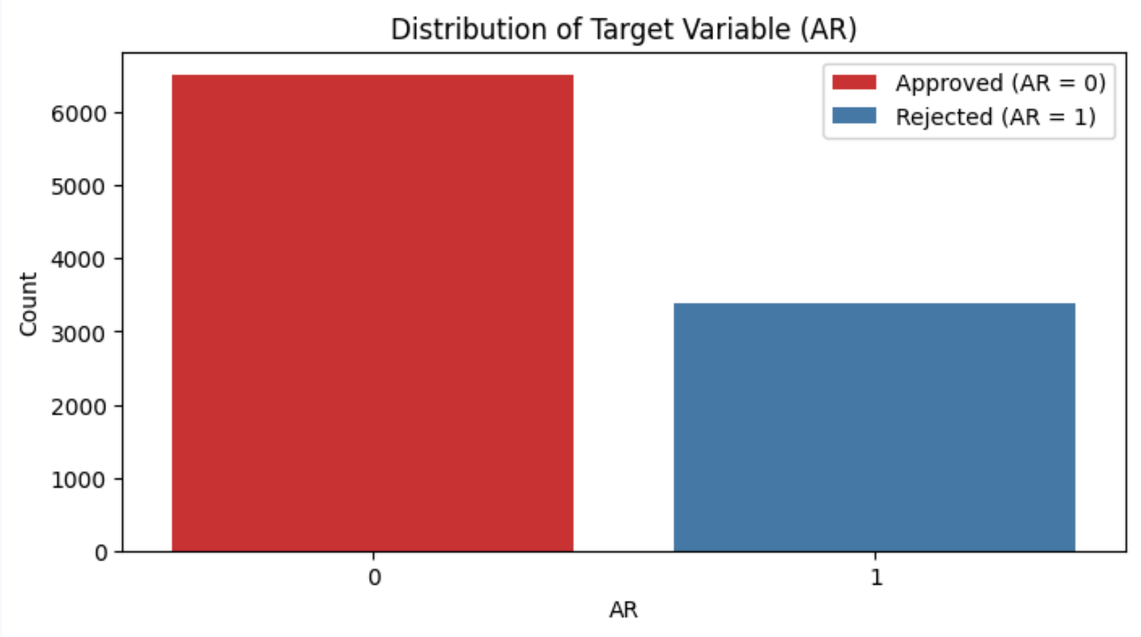
Dataset Description

DATASET

- Registry of loan applications for manual approval.
- Target variable: AR (Accepted/Rejected)
- Features: Various characteristics of loan applications (e.g., applicant income, credit score, etc.)

Moneza

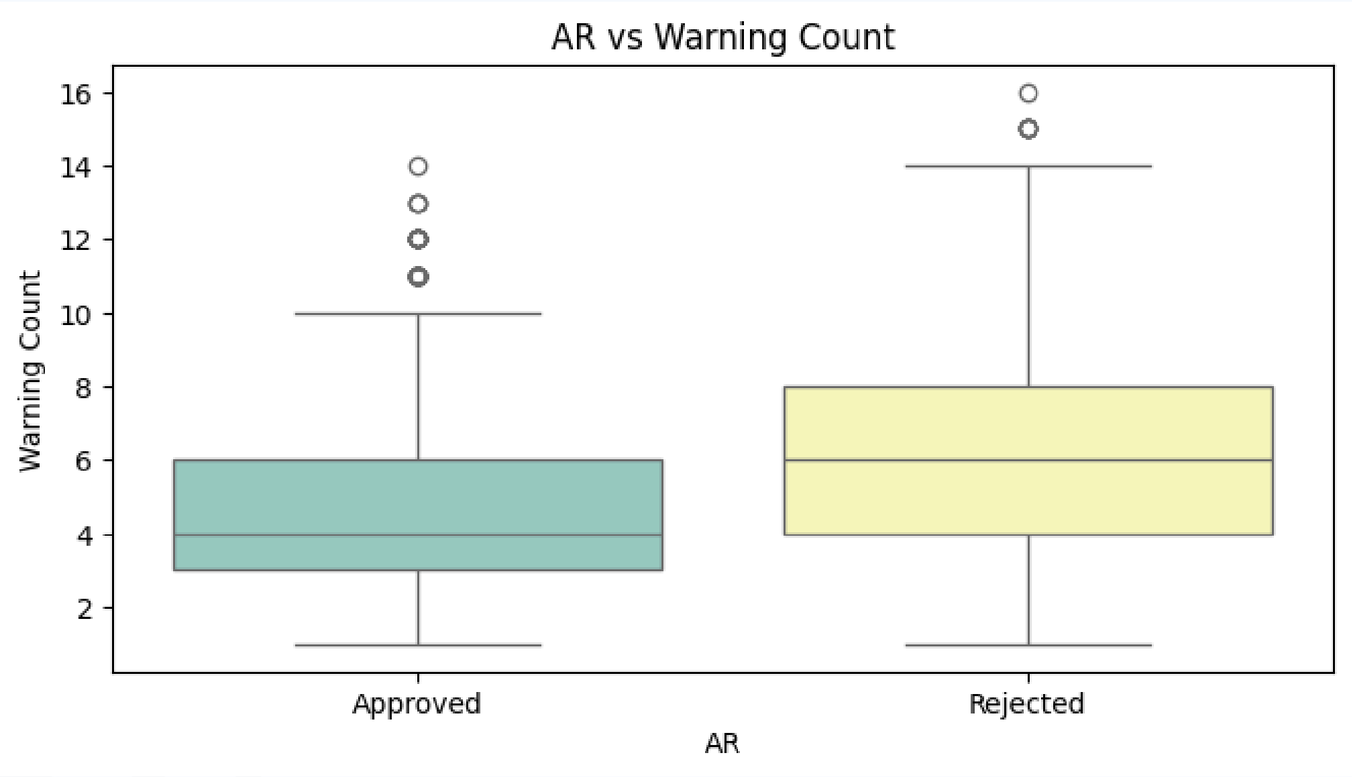
EDA
Distribution of Target Variable



The distribution shows how approved and rejected loan application are distributed in the dataset

Moneza

AR vs Warning Count (Numerical Analysis)



- These box plot gives insights to optimize the loan approval process, ensuring that decisions are fair, consistent, and aligned with risk management objectives.

i.e There is slighty rejected loans approval due to the number of warning counts

Moneza

Model Development

Logistic Regression

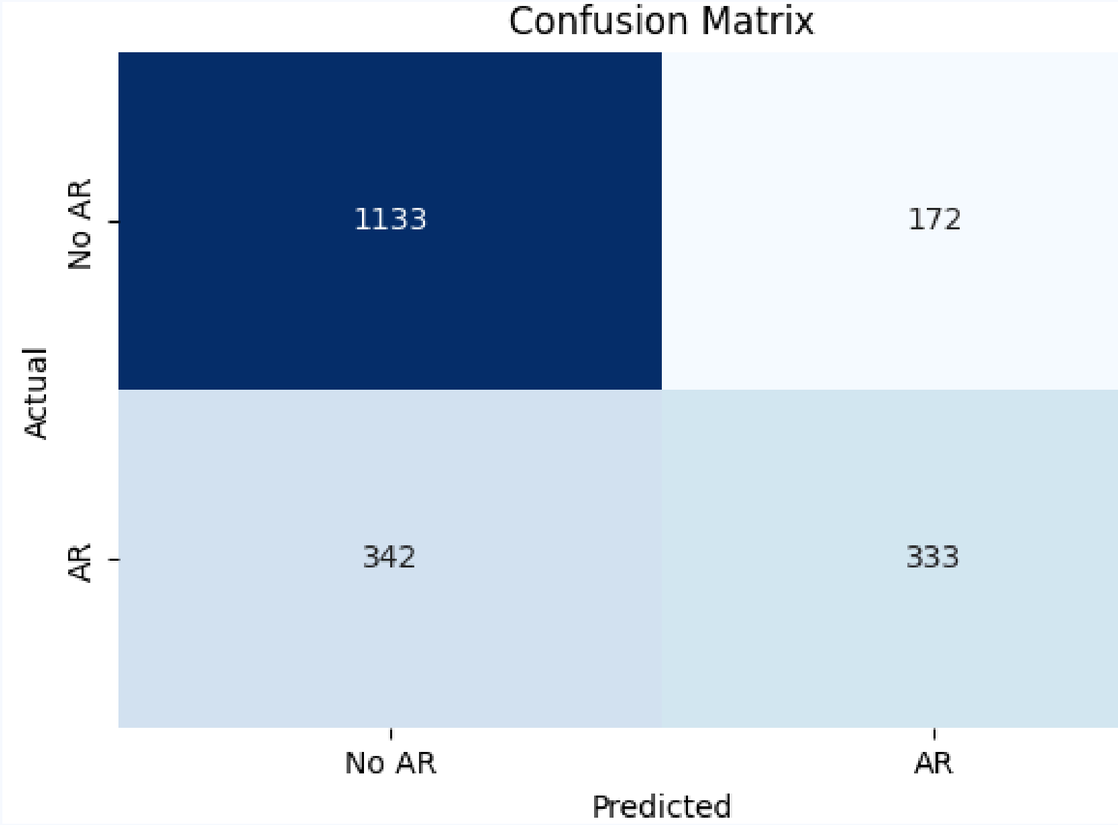
- Logistic Regression model used.
- Key features: Credit Card Check,Accounts Check, Warning ,client firtst loan etc
- Preprocessing: Encoding,Scaling, handling missing values, handling imbalance data
- I also used other models eg. Random Forest, SVM, XG BOOST and also Hyperparameter Tuning

Moneza

Model Performance

Confusion Matrix

- True Positive(T.P) - Correctly Approved (333) instances
- False Positive(F.P) - Incorrectly approved (172) instances
- True Negative(T.N) - Correctly Rejected (1133) instances
- False Negative (F.N) - Incorrectly Rejected (342) instances_



Moneza

Model Performance

Logistic Regression

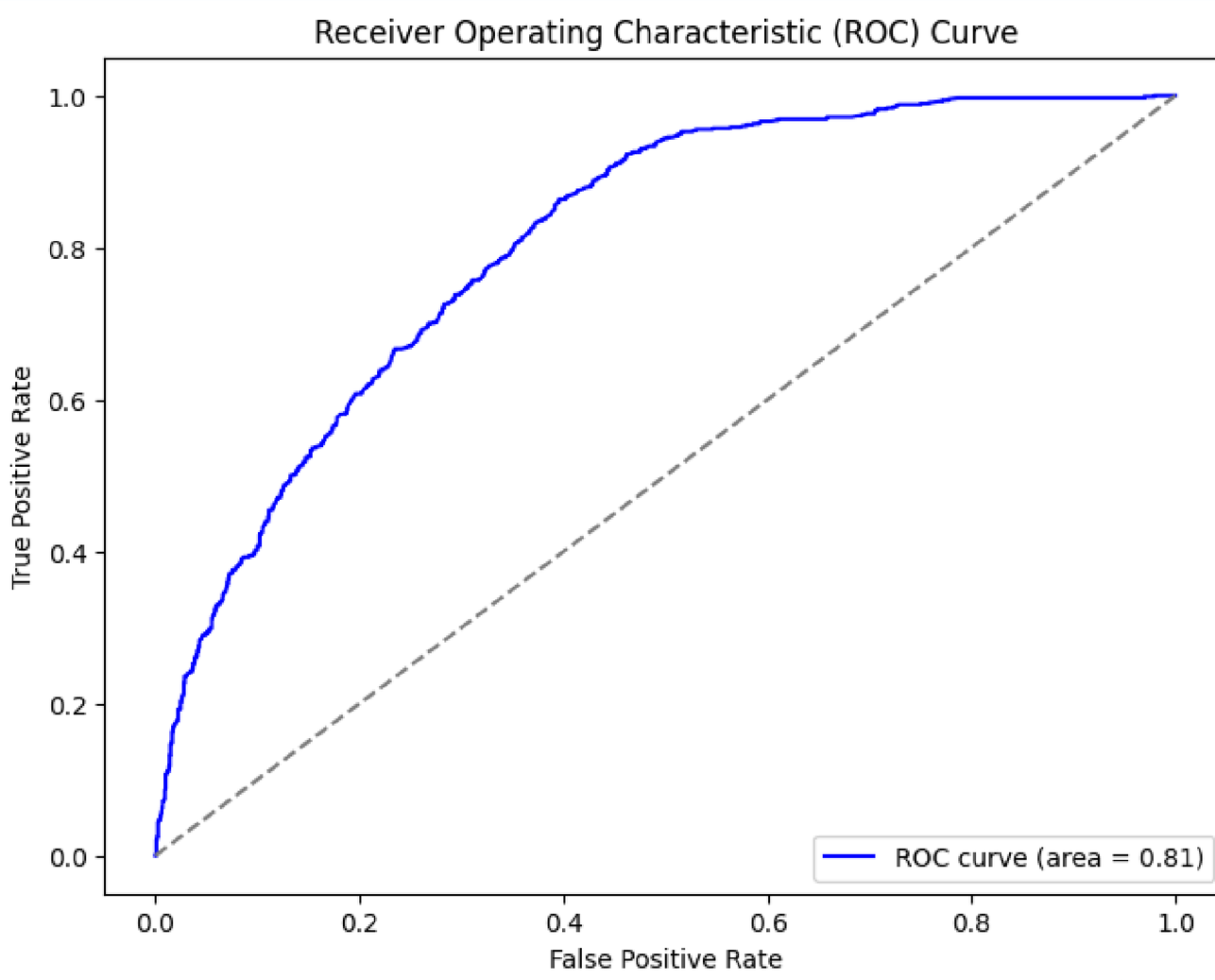
- Accuracy: Overall, the model predicted correctly 74.04% of all cases (accuracy score).
- Precision: Of all cases predicted as having AR, 66% actually had AR (precision for class 1).

Moneza

Model Performance

Logistic Regression

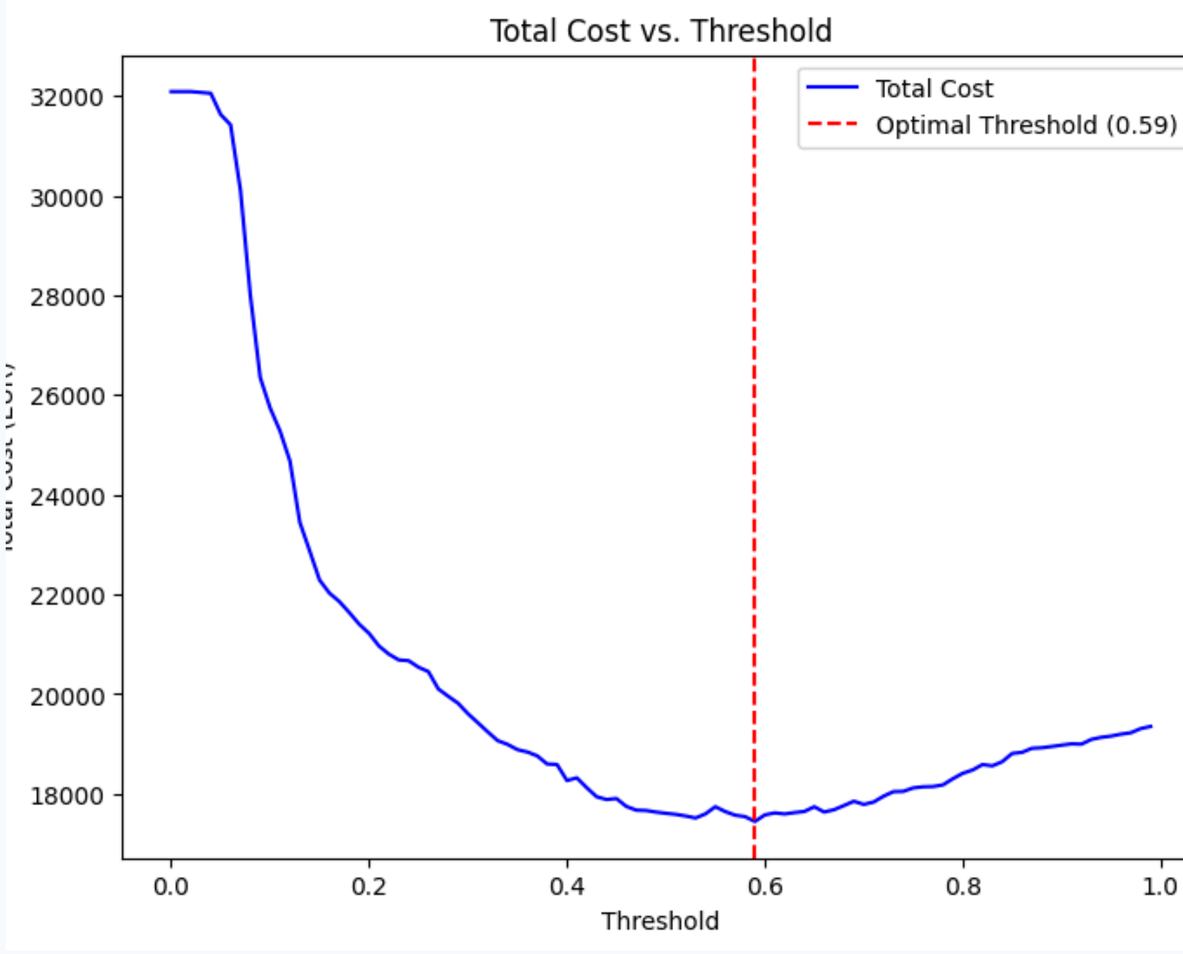
An AUC of 0.81 suggests that there is a 81% chance that the model will be able to distinguish between positive and negative classes.



Moneza

Optimized Threshold

The blue line shows the total cost at different threshold values. Initially, at lower thresholds, the cost is high due to many bad loans being approved. As the threshold increases, the cost decreases and reaches a minimum. Beyond this optimal point, the cost rises again as more good loans are rejected.



Moneza

Cost Comparison: Manual vs. Logistic Regression in Loan Approvals

- **Threshold Selection:**
- Threshold of 0.59 minimizes total cost.
- Balances trade-off between false positives and false negatives.
- **Cost Comparison:**
- Manual Processing Cost: 17,613 euros.
- Model-Based Decision Cost: 17,448 euros.
- Cost Reduction: 165 euros.
- **Conclusion:**
- Using the logistic regression model is more efficient.
- The model is also more economical for loan approvals.

Moneza

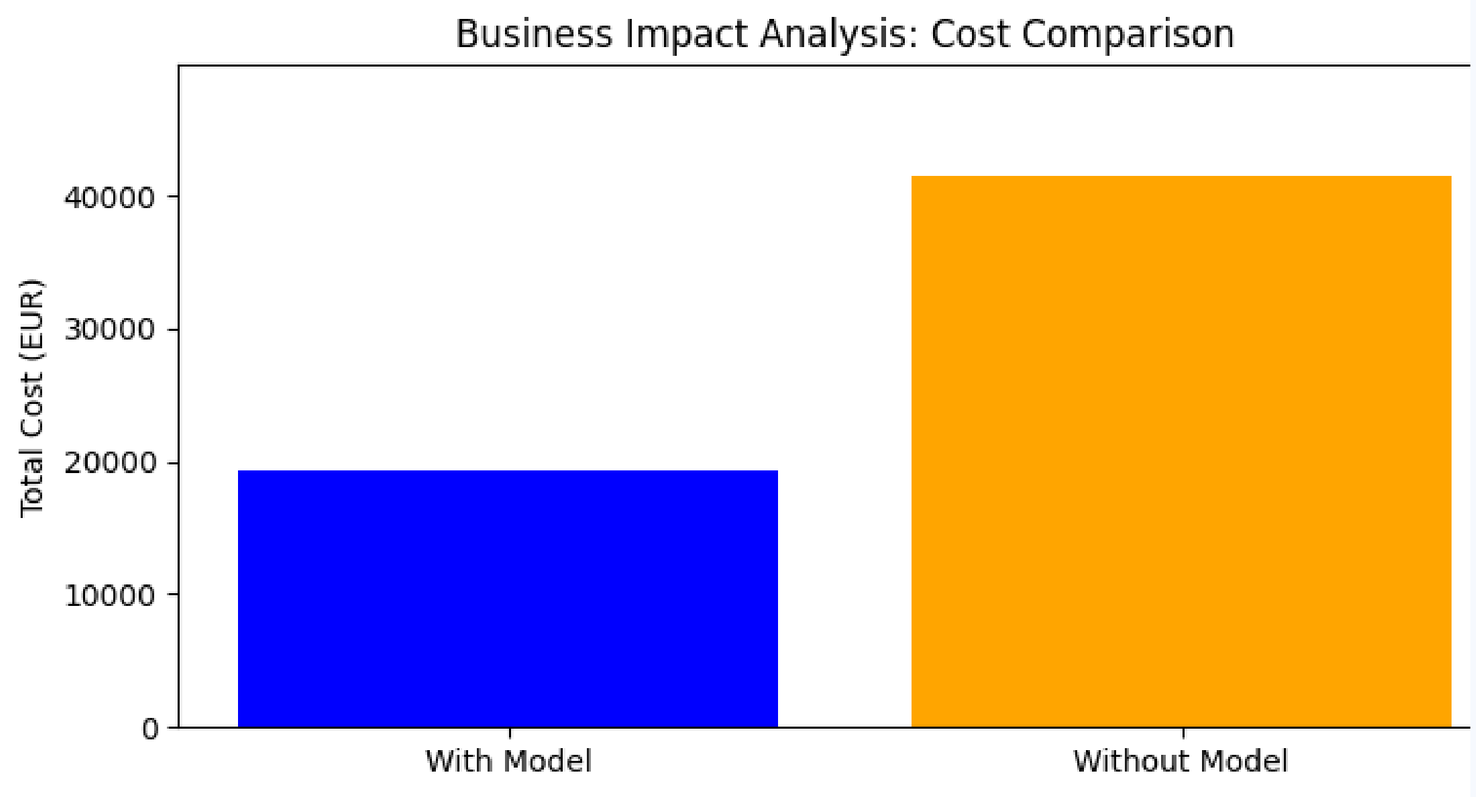
Business Impact

Bank Loan

- Significant cost saving
- More accurate loan approval decisions.
- Reduced financial losses from incorrect approvals/rejections.

Moneza

Business Impact Analysis



- With the Model: Significantly lower total costs due to reduced false approvals and rejections, leading to enhanced cost efficiency.
- Without the Model: Higher total costs from processing all applications without predictive filtering, indicating potential financial inefficiencies.

Moneza

Conclusion

- Logistic regression is the best model in predicting for loan approval in the given dataset.

- Achieved cost savings and improved decision accuracy.

Moneza

THANK YOU

Call : +254 700907204

kiptochmike33@gmail.com