Credit Card Default Prediction

Analyzing risks and utilizing machine learning techniques.

Understanding the Problem

Financial Burden

Credit card delinquency incurs significant costs annually, affecting businesses and consumers alike.

Model Limitations

Traditional prediction models often fail to accurately forecast credit defaults, leading to inefficiencies.

Machine Learning Solution

Our project aims to leverage machine learning to identify customers at high risk of defaulting.

Data Overview

Dataset Size

The dataset consists of 30,000 credit card records, providing a substantial basis for analysis.

Data Source

Records were collected from Taiwan in 2005, ensuring regional relevance in the analysis.

Payment History

Includes six months of payment history, important for tracking consumer behavior and trends.

Default Rate

A notable default rate of 22% emphasizes the necessity for predictive modeling.

Key Features Identified

Delinquency Flags

These flags indicate late payments which are crucial for identifying risk factors.

Payment-to-Bill Ratios

This ratio helps in assessing how well customers manage their payments relative to their bills.

Credit Utilization Metrics

Understanding credit utilization is key in predicting customer defaults and financial health.

Machine Learning Models Used

Overview of the Models

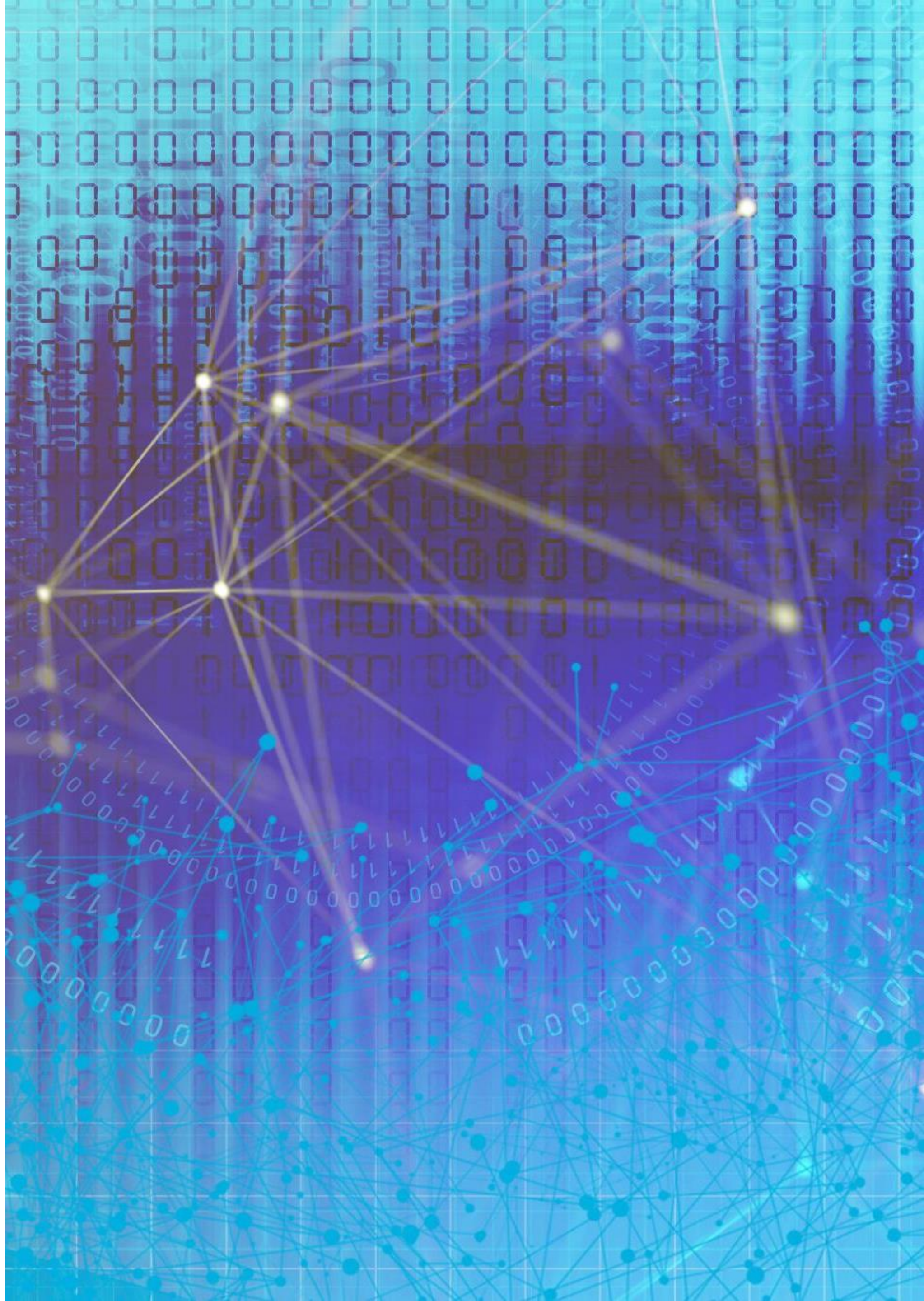
We utilized three models: Logistic Regression, SVM, and MLP Neural Net. Each has unique strengths for predictive analysis.

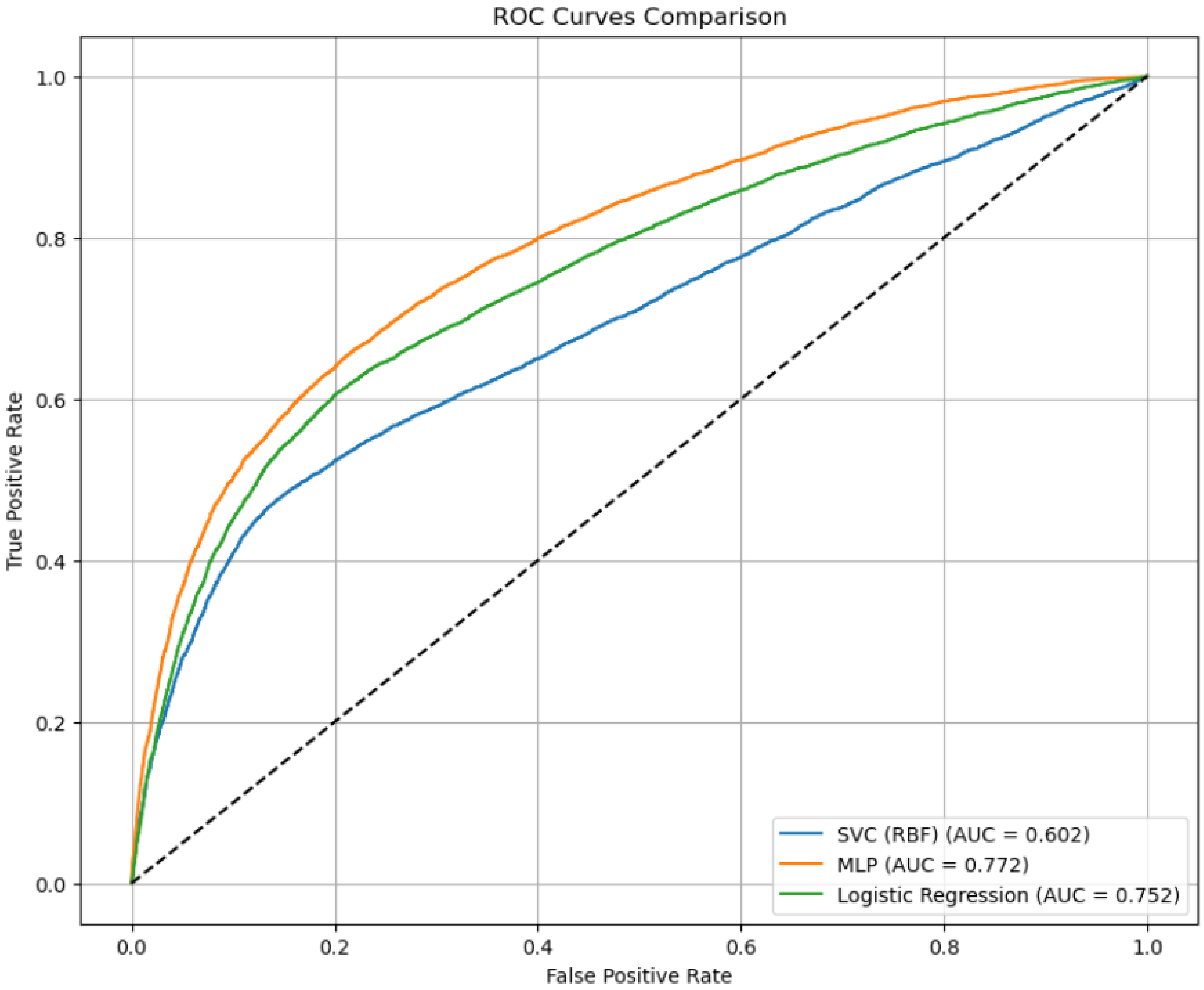
Performance of Logistic Regression

The Logistic Regression model achieved an AUC score of 0.75, indicating decent predictive ability.

Performance of MLP Neural Net

The MLP Neural Net outperformed with an AUC of 0.77, demonstrating superior predictive power.




ROC Curves Comparison
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
SVC (RBF) (AUC = 0.602)
MLP (AUC = 0.772)
Logistic Regression (AUC = 0.752)

Modeling Process Explained

Data Scaling Importance

Data scaling ensures the features contribute equally to the analysis. This is crucial for effective modeling.

Principal Component Analysis

PCA helps in reducing dimensionality while retaining 90% variance, enhancing model efficiency.

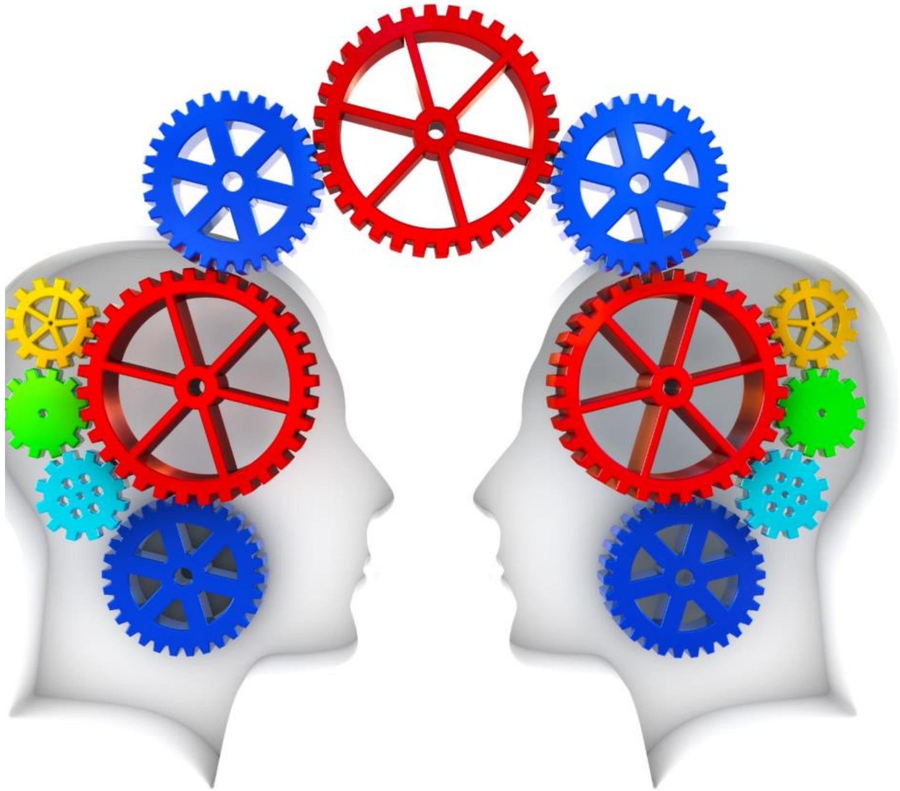
Cross-Validation Technique

5-fold cross-validation helps in assessing model performance and reducing overfitting.

Hyperparameter Tuning

Optimizing hyperparameters is essential to improve the model's predictive accuracy.

Results Achieved



Prediction Accuracy

The model achieved a prediction accuracy of 77.2%, significantly outperforming random guessing.

Identification Rate

The model identifies 4 out of 5 true defaulters, showcasing its effectiveness.

Real-World Reliability

This model proves reliable in real-world scenarios, enhancing decision-making processes.

Business Value of the Model

Proactive Risk Management

Identifying risky customers early helps in mitigating financial risks effectively. This proactive approach safeguards business finances.

Adjusting Credit Limits

Adjusting credit limits based on risk assessment prevents potential losses. This ensures financial stability for the business.

Improving Customer Relations

A better understanding of customer risk leads to improved relationships. Customers appreciate thoughtful credit management.

Next Steps in Development

Deeper Neural Networks

We aim to explore advanced architectures for improved model performance and accuracy.

Temporal Trend Features

Incorporating temporal data will enhance our predictions by capturing time-related patterns.

Explainable AI Techniques

We will implement methods that provide insights into model decisions, fostering trust and transparency.

Conclusion and Summary

MLP Model Performance

The MLP model achieved an AUC of 0.77, outperforming other models. This demonstrates its effectiveness in the task.

Feature Engineering Impact

Effective feature engineering played a crucial role in improving our results significantly. It enhanced the model's predictive capabilities.

Production-Ready System

We have a production-ready system for identifying high-risk credit card users. This system is prepared for real-world application.