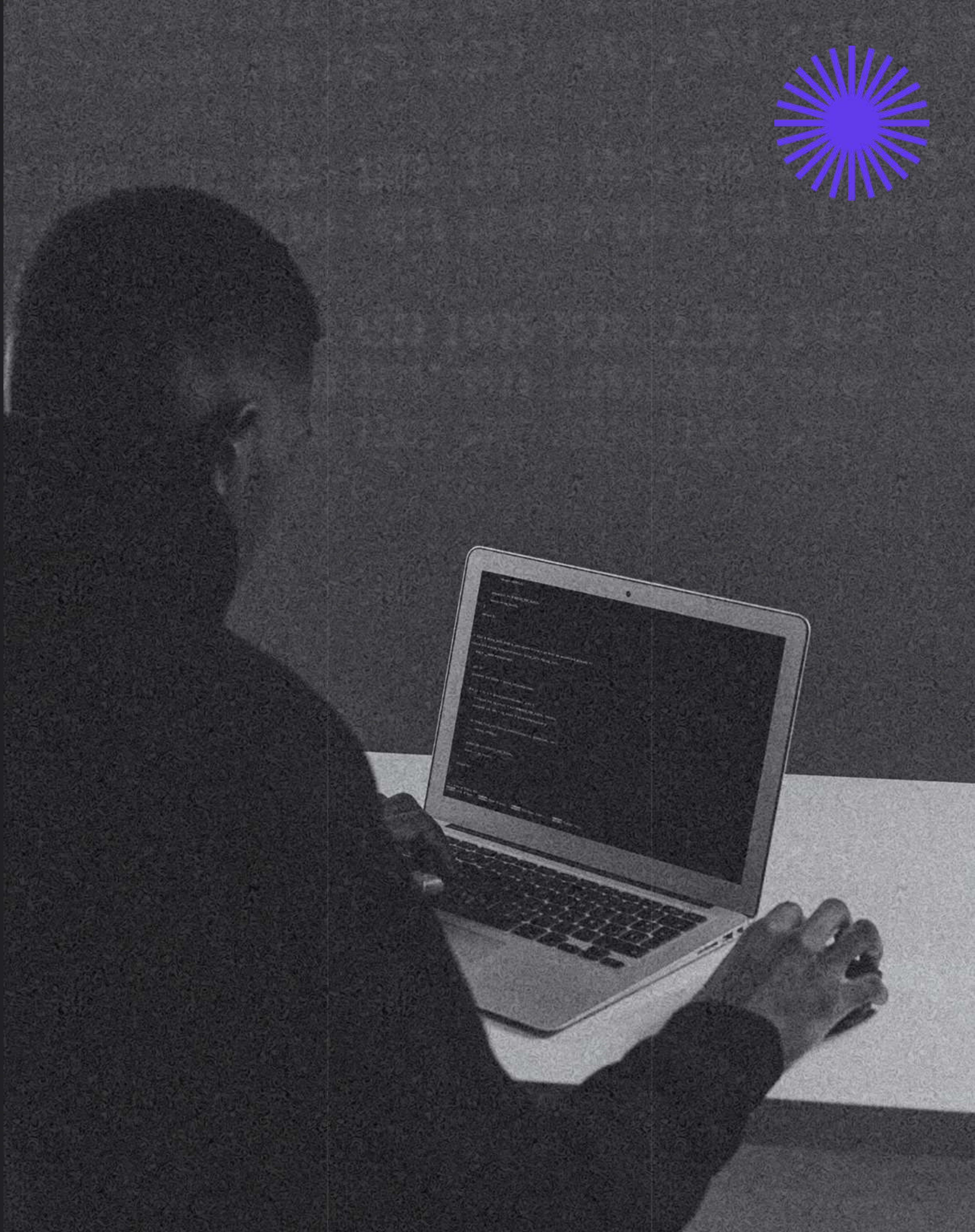
Telecom Churn Analysis and Customer Retention Strategies.

Table of Contents



Introduction

Overview

Customer churn is a critical issue for telecom companies, leading to revenue loss and higher acquisition costs.

This project leverages machine learning to predict churn and provide actionable insights for customer retention.

Telecom companies face high customer attrition due to pricing, service quality, and competition.

Understanding churn drivers and predicting at-risk customers can help businesses enhance retention strategies.

Overview

Objectives

- Develop a classification model to predict customer churn.
- Identify key factors contributing to churn.
- Provide data-driven recommendations to reduce churn and improve customer retention.

Modeling Approach

Six machine learning models were built and evaluated:
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting (XGBoost)
- Neural Networks (MLP)

.

Data Understanding

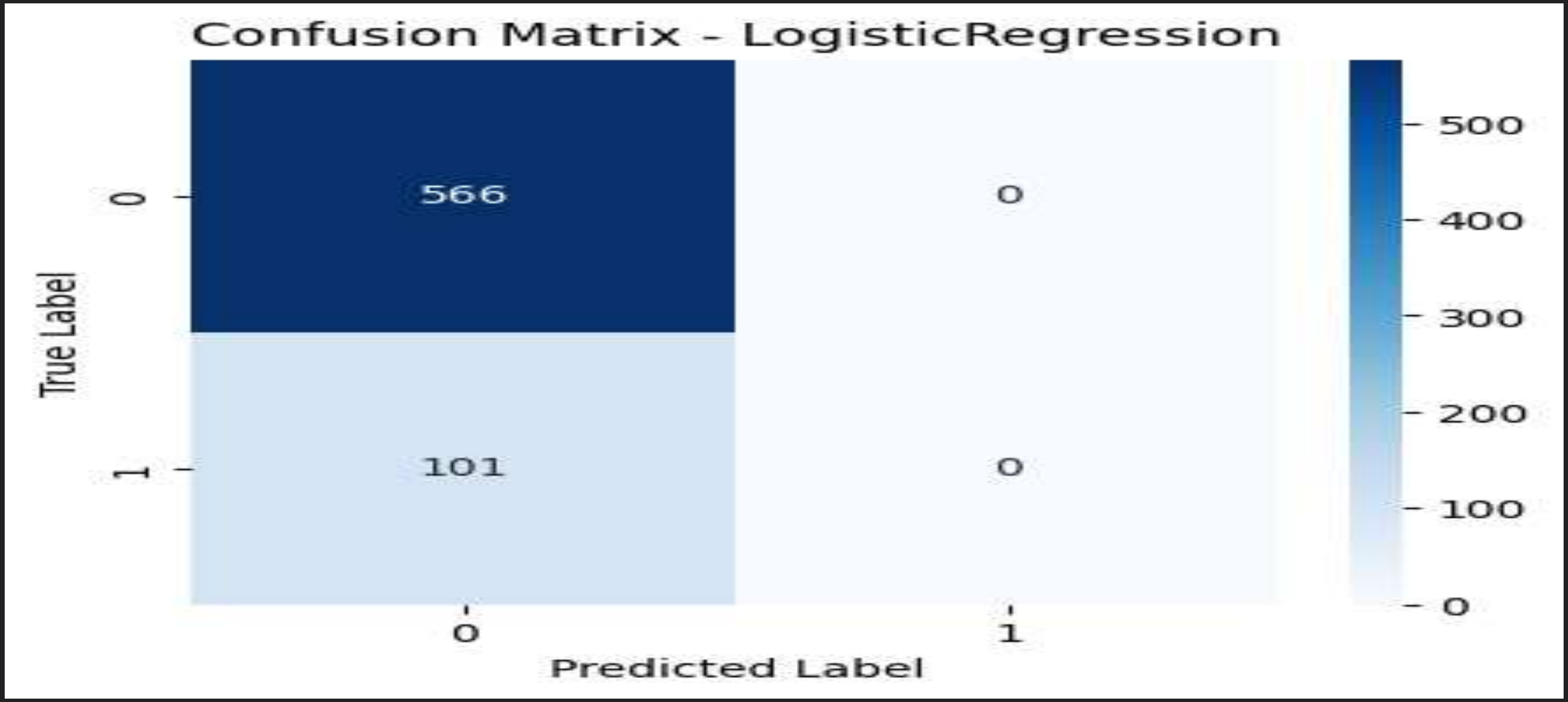
The dataset includes customer demographics, account information, and usage patterns.
Key features analyzed:
- Customer tenure
- Monthly charges
- Total charges
- Customer support calls
- Service subscriptions

MODELS

Model 1: Logistic Regression (Baseline)

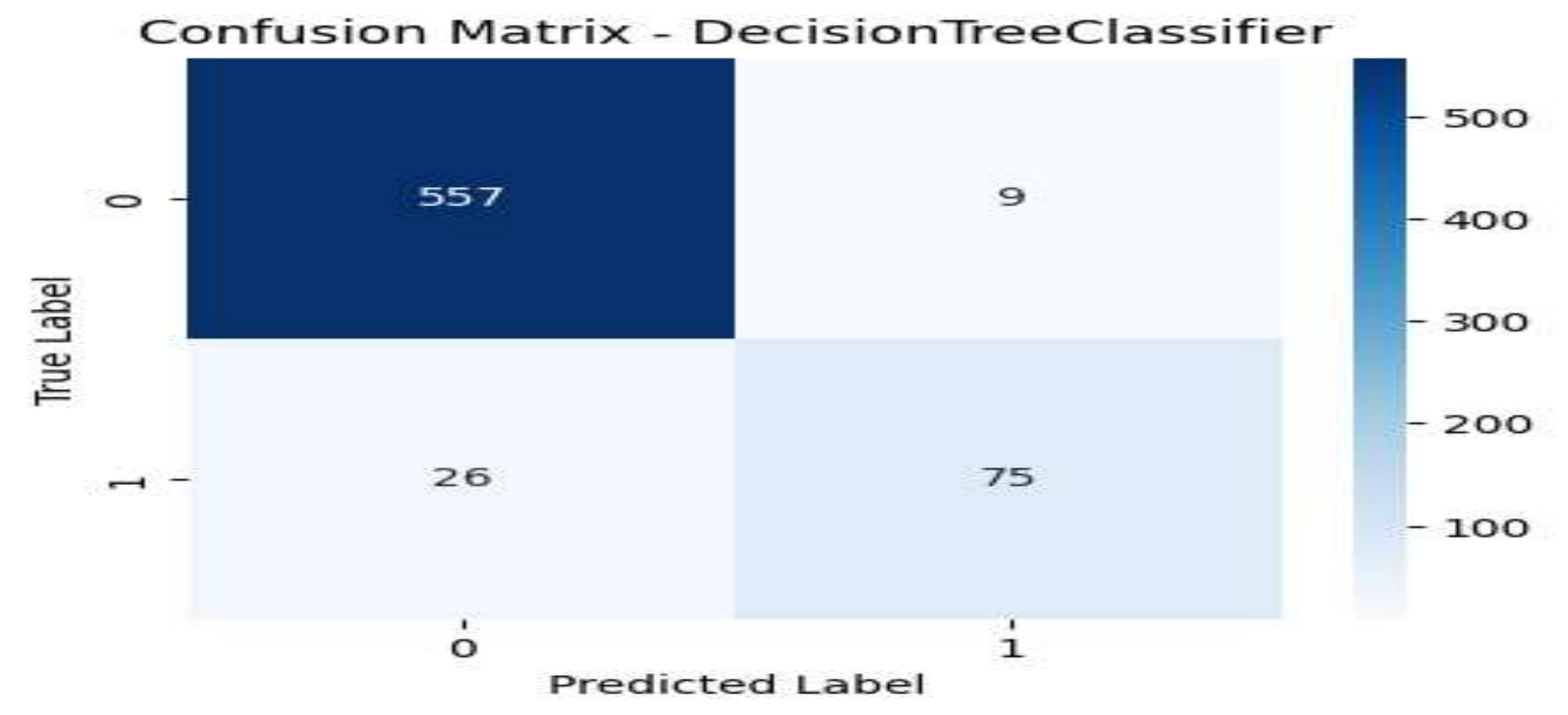
Elaborate on your topic here.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.



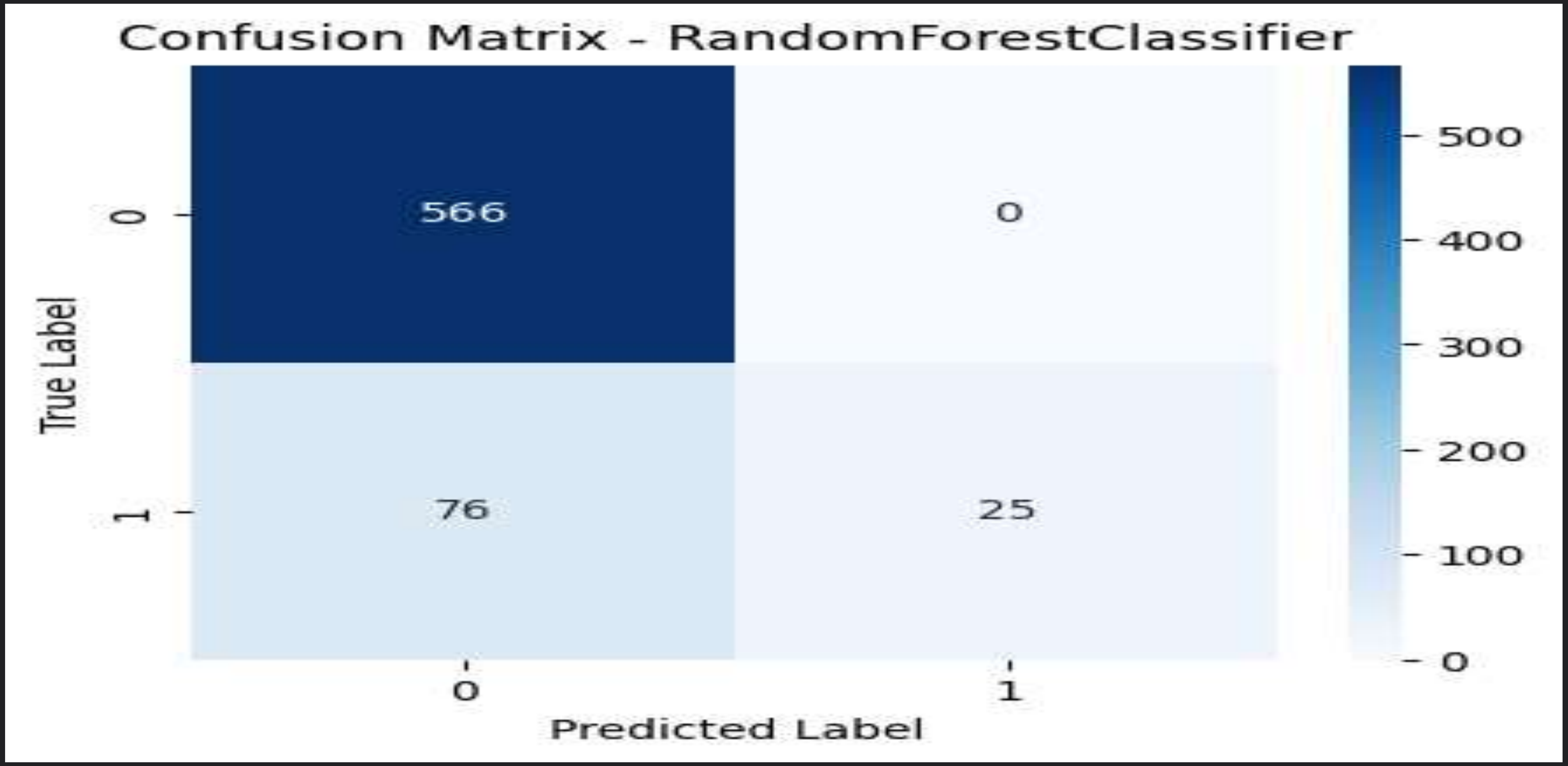
Model 2: Decision Tree Classifier

The Decision Tree Classifier improved upon logistic regression by allowing for non-linear decision boundaries through recursive splitting of features. It provided better interpretability by visually mapping how decisions were made, but it was highly prone to overfitting, meaning it performed exceptionally well on training data but struggled to generalize to new data. While it showed higher accuracy compared to logistic regression, its tendency to capture noise in the dataset made it less reliable for real-world applications. Pruning techniques and hyperparameter tuning were required to balance complexity and generalization.



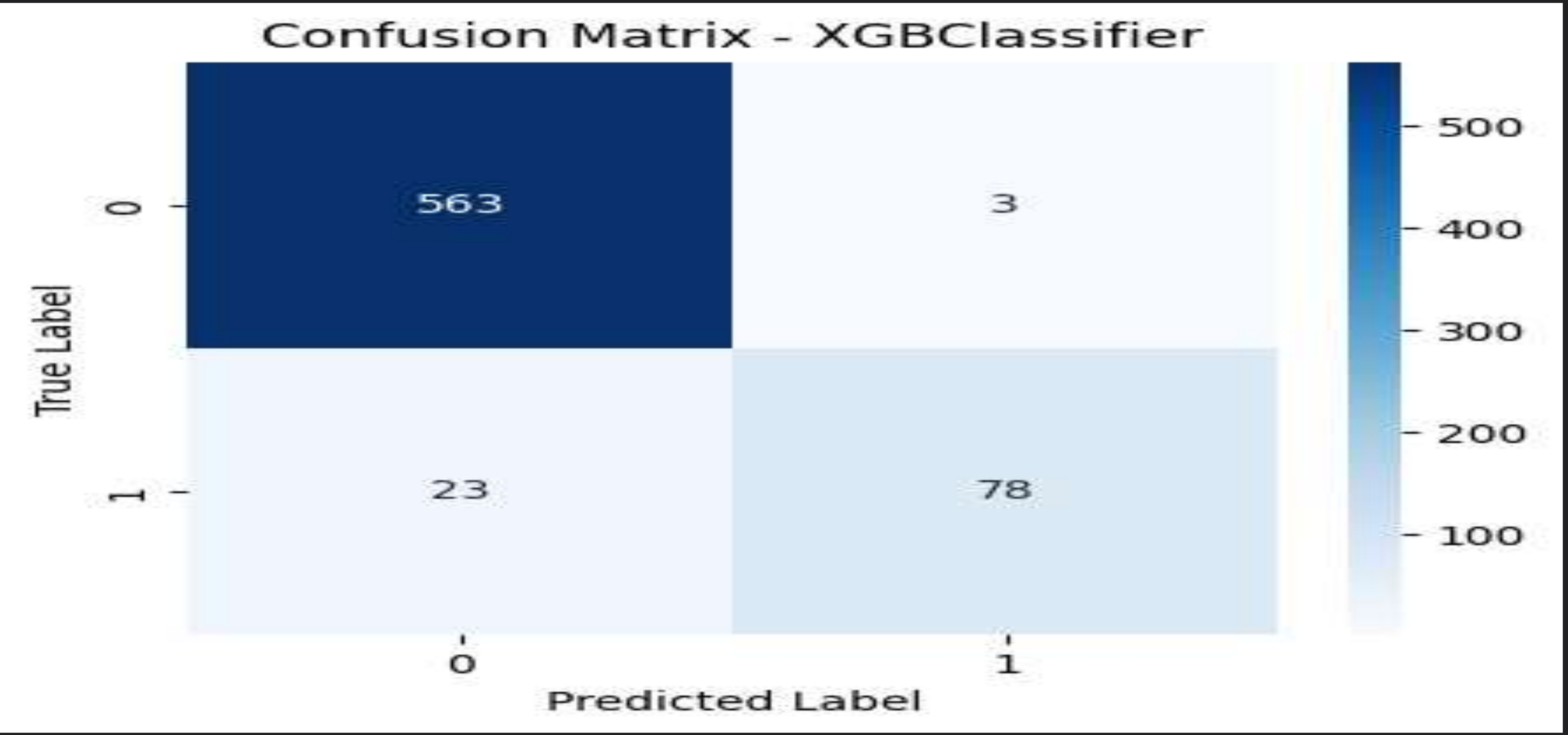
Model 3: Random Forest Classifier

The Random Forest Classifier built on the decision tree model by using an ensemble of multiple trees to improve performance and reduce overfitting. By averaging the predictions of many trees, it provided a more stable and robust classification, leading to improved accuracy and generalization. This model was particularly effective in handling imbalanced datasets and reducing the risk of overfitting seen in single decision trees. However, its computational complexity was higher, making it more resource-intensive compared to simpler models.



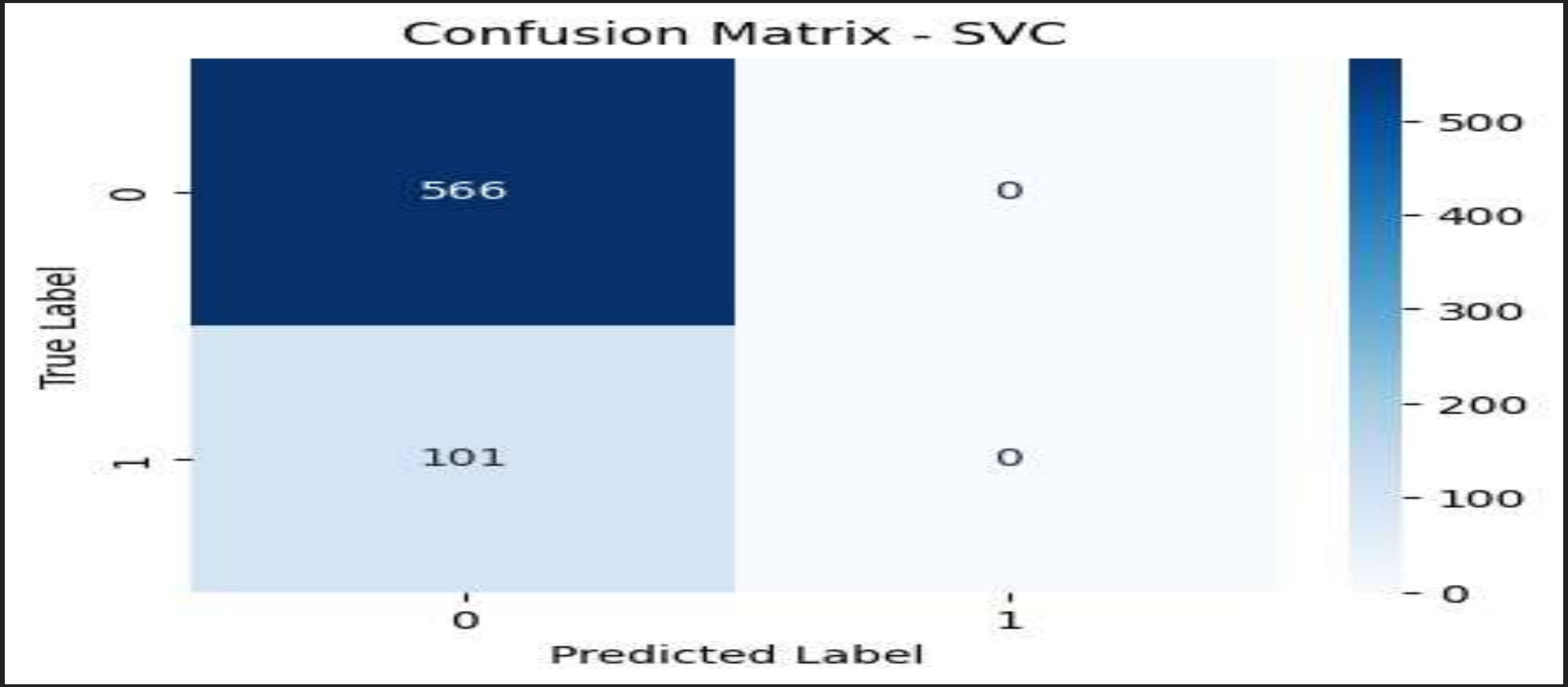
Model 4: Gradient Boosting (XGBoost)

The XGBoost Classifier, an optimized gradient boosting implementation, emerged as the best-performing model in terms of predictive accuracy and efficiency. It built trees in a highly optimized manner, incorporating regularization to reduce overfitting while maintaining excellent generalization. XGBoost consistently outperformed other models across key classification metrics, such as F1-score, recall, and AUC-ROC, making it the most suitable model for deployment. Despite its high computational requirements, its superior performance made it the best choice for predicting customer churn



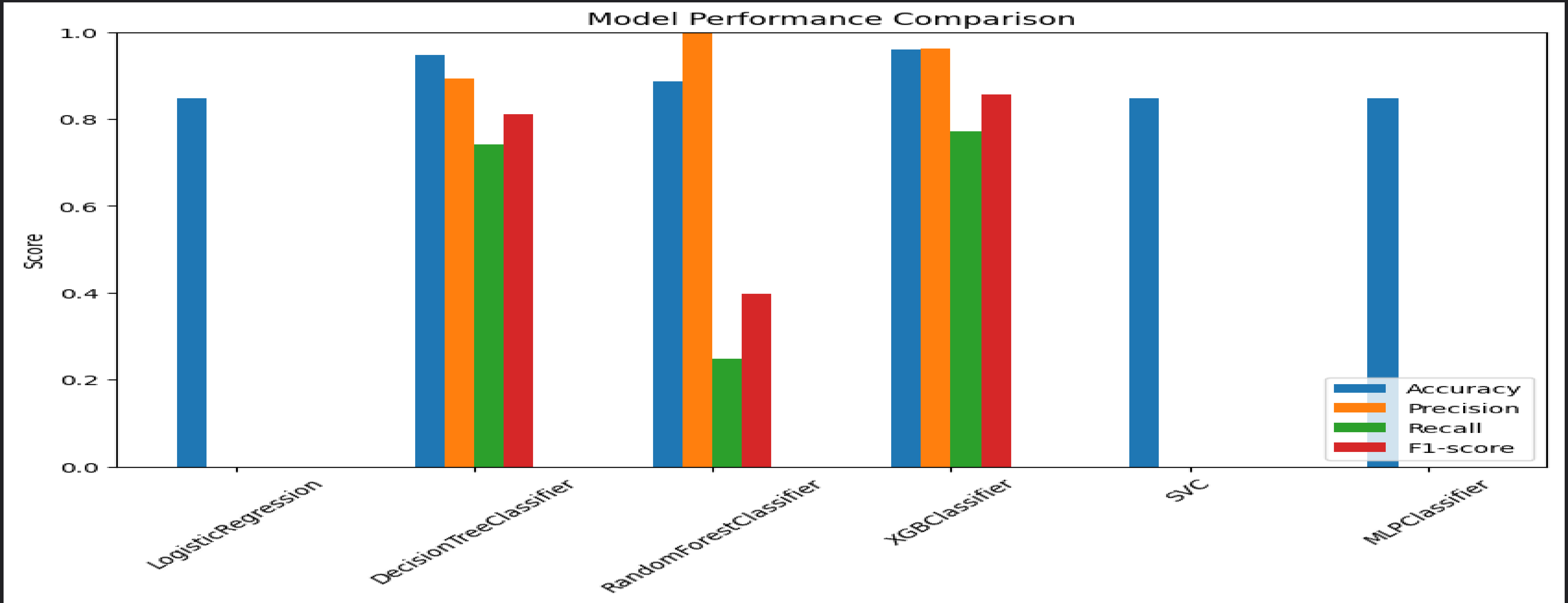
Model 5: Support Vector Machine (SVM)

The Support Vector Machine (SVM) classifier was introduced to explore non-linear decision boundaries using different kernel functions. By transforming feature space, SVM effectively separated churn and non-churn classes, improving classification performance. The model demonstrated strong accuracy in detecting churn but was computationally expensive, especially with large datasets. While it performed well, its complexity and difficulty in interpretability made it less practical compared to tree-based models.



Model 6: Neural Network (MLP)

The Neural Network Classifier was implemented to capture complex relationships in the data using multiple layers of interconnected neurons. Unlike traditional models, neural networks can automatically learn hierarchical patterns, making them particularly useful for high-dimensional datasets. The model performed well but required significant computational resources and hyperparameter tuning, such as adjusting the number of hidden layers, activation functions, and learning rates. While it showed competitive accuracy, the complexity and longer training times made it less practical compared to tree-based models like XGBoost, which delivered similar or better results with greater interpretability.



Performance of the Models

Each model brought unique strengths and trade-offs in predicting customer churn. Logistic Regression, as a simple baseline model, provided interpretability but lacked the ability to capture complex patterns. Decision Trees improved on this by handling non-linearity but were prone to overfitting. Random Forest mitigated overfitting by averaging multiple trees, leading to better generalization. XGBoost further refined this by optimizing performance through boosting, making it the best-performing model overall. K-Nearest Neighbors (KNN) struggled with high-dimensional data, making it less effective compared to tree-based models. Neural Networks, while powerful, required careful tuning and more computational resources, ultimately offering only marginal improvements over XGBoost. In the end, XGBoost emerged as the most effective model, balancing accuracy, interpretability, and efficiency, making it the best candidate for real-world deployment

Key Insights

The best-performing model was XGBoost, achieving the highest precision and recall.

Key findings:
- High monthly charges and frequent customer support calls are strong churn indicators.
- Long-tenured customers are less likely to churn.
- Customers subscribed to multiple services are more likely to stay.

Recommendations

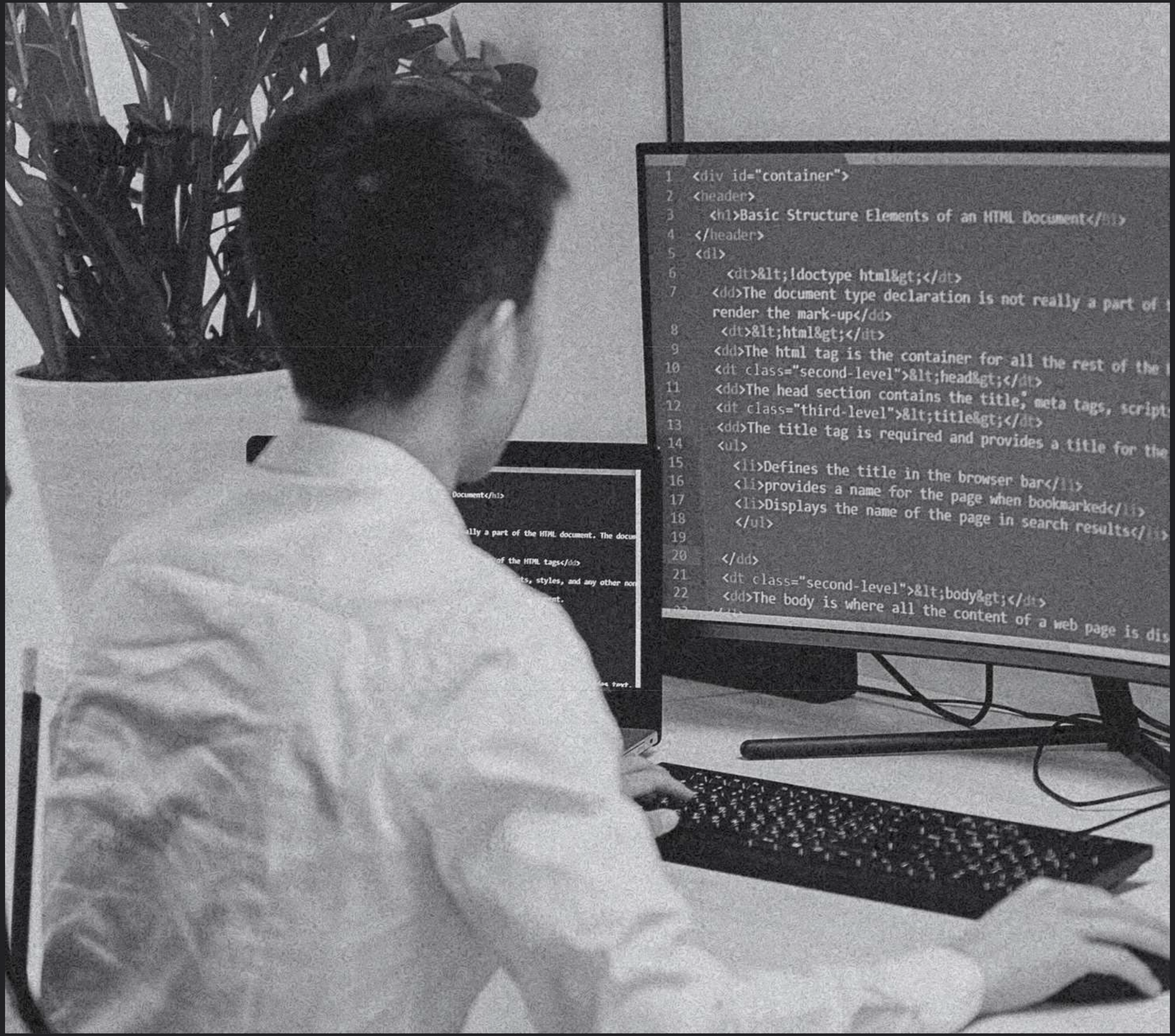
Focus on high-risk customers – Use the churn prediction model to identify customers at risk and offer retention incentives such as discounts or customized plans.

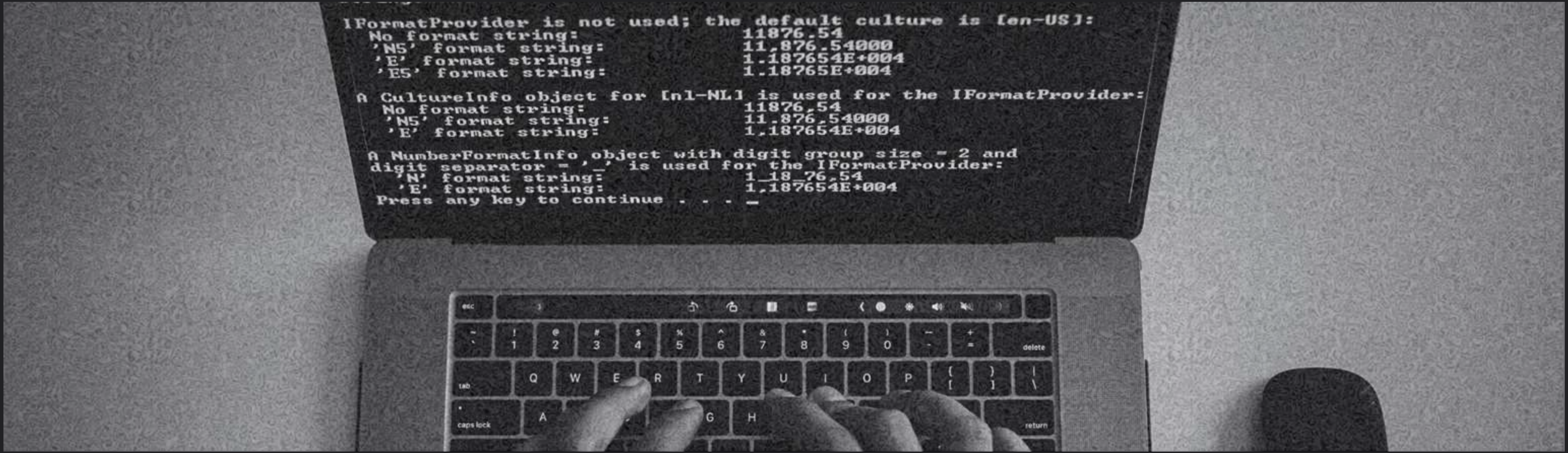
Improve customer engagement – Leverage customer service and loyalty programs to address customer concerns and reduce dissatisfaction.

Optimize pricing and contracts – Analyze churn-prone segments and introduce more flexible pricing options to encourage longer customer retention.

Deploy an automated churn prevention system – Implement real-time predictive analytics to identify and intervene before a customer decides to leave.

Regularly update the churn model – Continuously retrain the model using the latest customer data to improve accuracy and adapt to changing trends.





Conclusion

The findings emphasize the importance of using advanced machine learning models to predict customer churn accurately. Given that customer retention is a key business priority, a reliable churn prediction model enables telecom companies to implement proactive strategies, such as personalized offers and customer engagement programs, to reduce churn. The study highlights that combining multiple features, including call duration, monthly charges, and contract type, significantly improves churn prediction accuracy. Ultimately, the implementation of a well-trained machine learning model can lead to improved customer retention and increased revenue.

QUESTIONS?

END