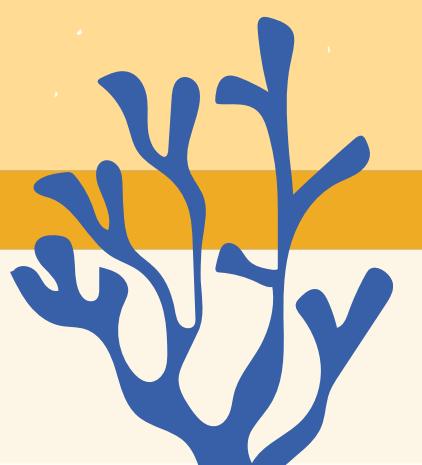




TREEPEDIA

CUSTOMER CHURN PREDICTION

A Data-Driven Approach to Retain
Customers and Boost Profitability



AGENDA

Introduction

Business Challenge

Data-Driven Approach

Data Insights

Machine Learning Model

Model Evaluation

Conclusion



TREEEPEDIA

► a leading e-commerce company, faces high customer churn rates of 16.35% (531 out of 3248 customers), impacting revenue and growth.



BUSINESS CHALLENGE

UNDERSTANDING CHURN

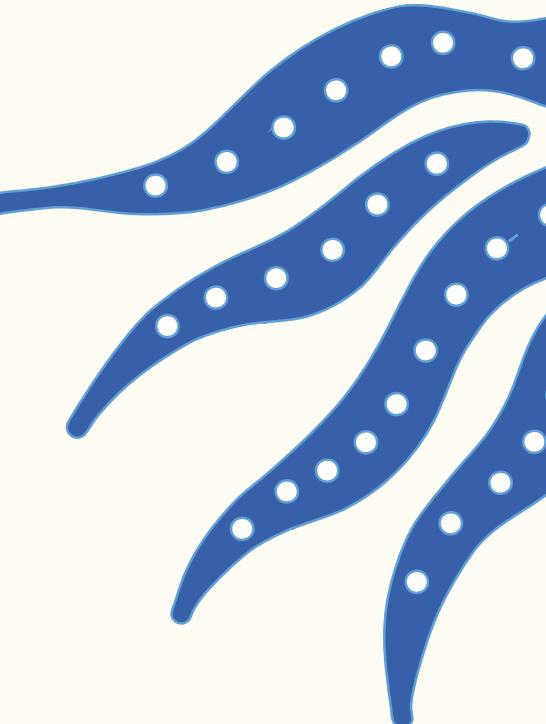


💡 Churn is the cessation of purchasing activity or engagement with the platform.



Customer churn is costly, representing missed purchases, reduced customer lifetime value, and negative brand impact.

💡 Understanding churn drivers is crucial for targeted interventions to improve customer satisfaction, loyalty, and profitability



OBJECTIVES



Develop a predictive model to accurately identify customers at high risk of churn.



Optimize resource allocation for retention efforts by minimizing false positive predictions.

DATA-DRIVEN APPROACH



1. Data Collection and Preparation



3941 initial records,
3248 after cleaning

3. Model Development

Train and evaluate
machine learning models



2. Exploratory Data Analysis (EDA)



Uncover patterns
and correlations in
customer behavior

4. Deployment and Monitoring



Deploy the model and
continuously monitor its
performance

DATA INSIGHTS

Uncovering Churn Patterns



PRODUCT PREFERENCES

- Laptop & Accessory category has the lowest churn (4.18%).
- Mobile Phone category has the highest churn (23.51%).

MARITAL STATUS

Single customers churn more often (20.15%) than married customers (14.51%)

OTHER FACTORS

Tenure, satisfaction scores, days since last order, and cashback amounts also influence churn



MACHINE LEARNING MODEL



FINANCIAL IMPLICATIONS

ASSUMPTIONS

Average Customer Lifetime Value (CLTV): \$500

Marketing Budget: 13.6% of retained customer revenue

Marketing Cost per Customer: \$68

FINANCIAL IMPACT OF CHURN

Lost Revenue: \$500 per churned customer

Wasted Resources: \$68 per customer incorrectly predicted to churn

POTENTIAL FOR MODEL

Accurately identify and retain high-risk customers

Avoid unnecessary marketing spend on loyal customers

Benchmark Model

31.631% profit lift

Demonstrates potential to significantly increase profit compared to no model

Adresses Class Imbalance

More retained customers than churned

XGBoost with Random Over Sampler

The model shows signs of overfitting, indicating a need for further optimization



This presentation is optimized for view on a large screen.

Final Model

Reduced total loss by
69.894 (\$37,044)

Increased profit by
31.30% (\$68,544)

Compared without
machine learning



Optimized XGBoost

Hyperparameter tuning further
improved XGBoost's performance and
addressed the overfitting issue

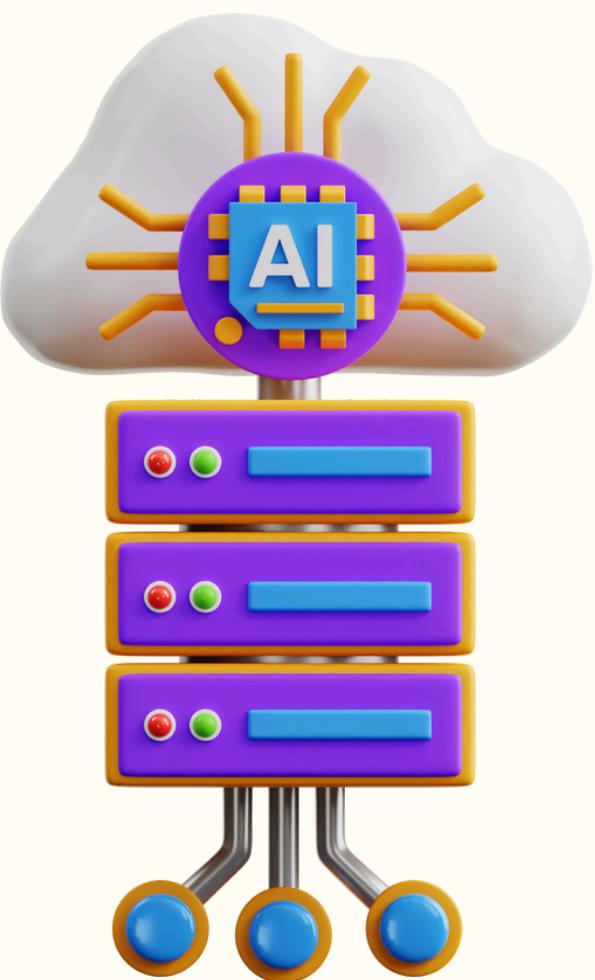
Best Performance

93% accuracy, 85%
recall, and 77%
precision

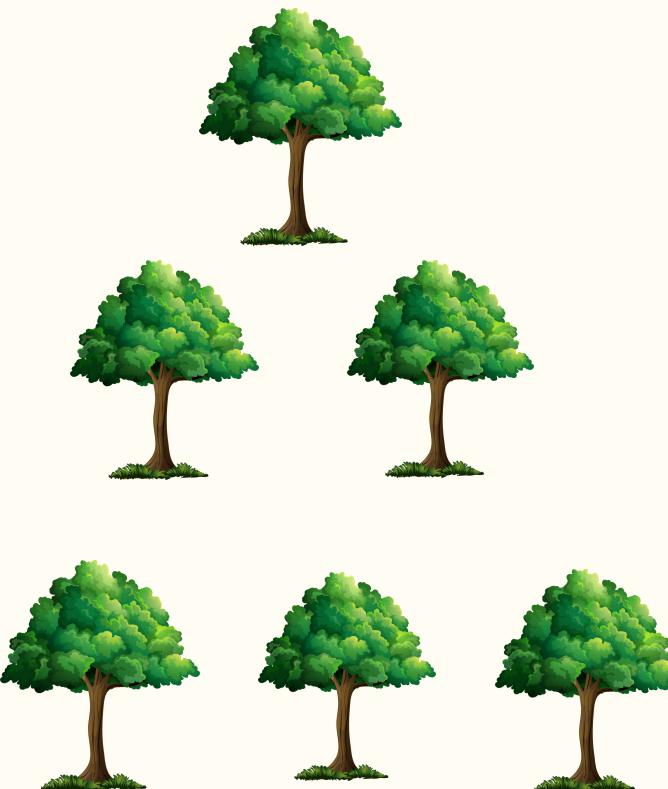


This presentation is optimized for whiteboard use

HOW XGBOOST WORKS?



XGBoost is an ensemble learning method that combines multiple decision trees to make predictions



Each tree is trained to correct the errors of the previous trees.

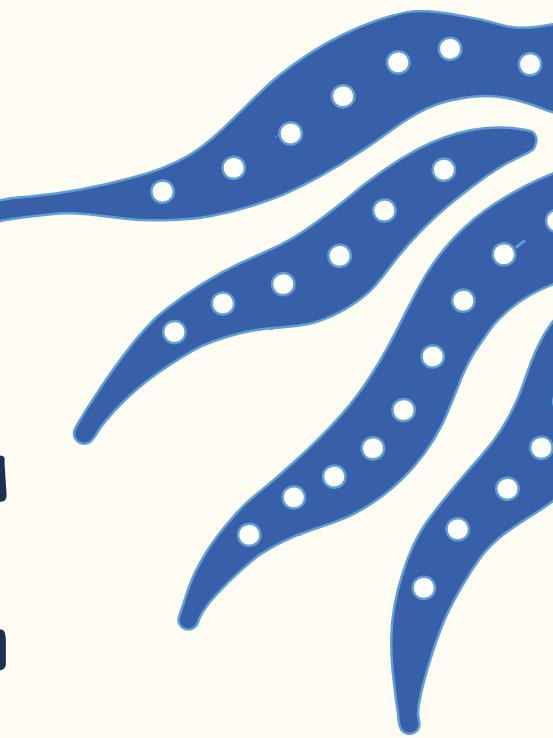


The final prediction is a weighted average of the predictions of all the trees

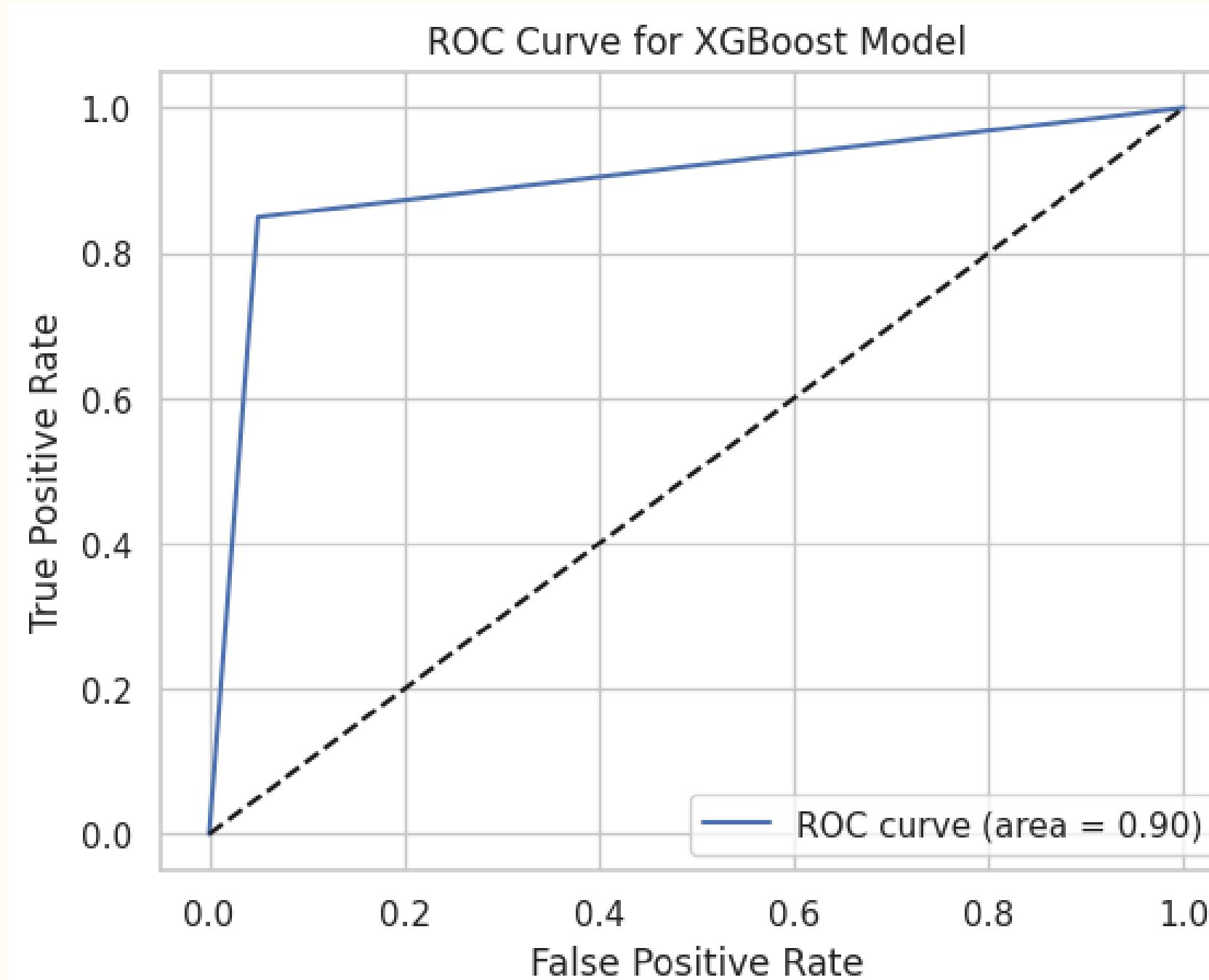
XGBoost is known for its speed, accuracy, and scalability

MODEL EVALUATION





ROC CURVE



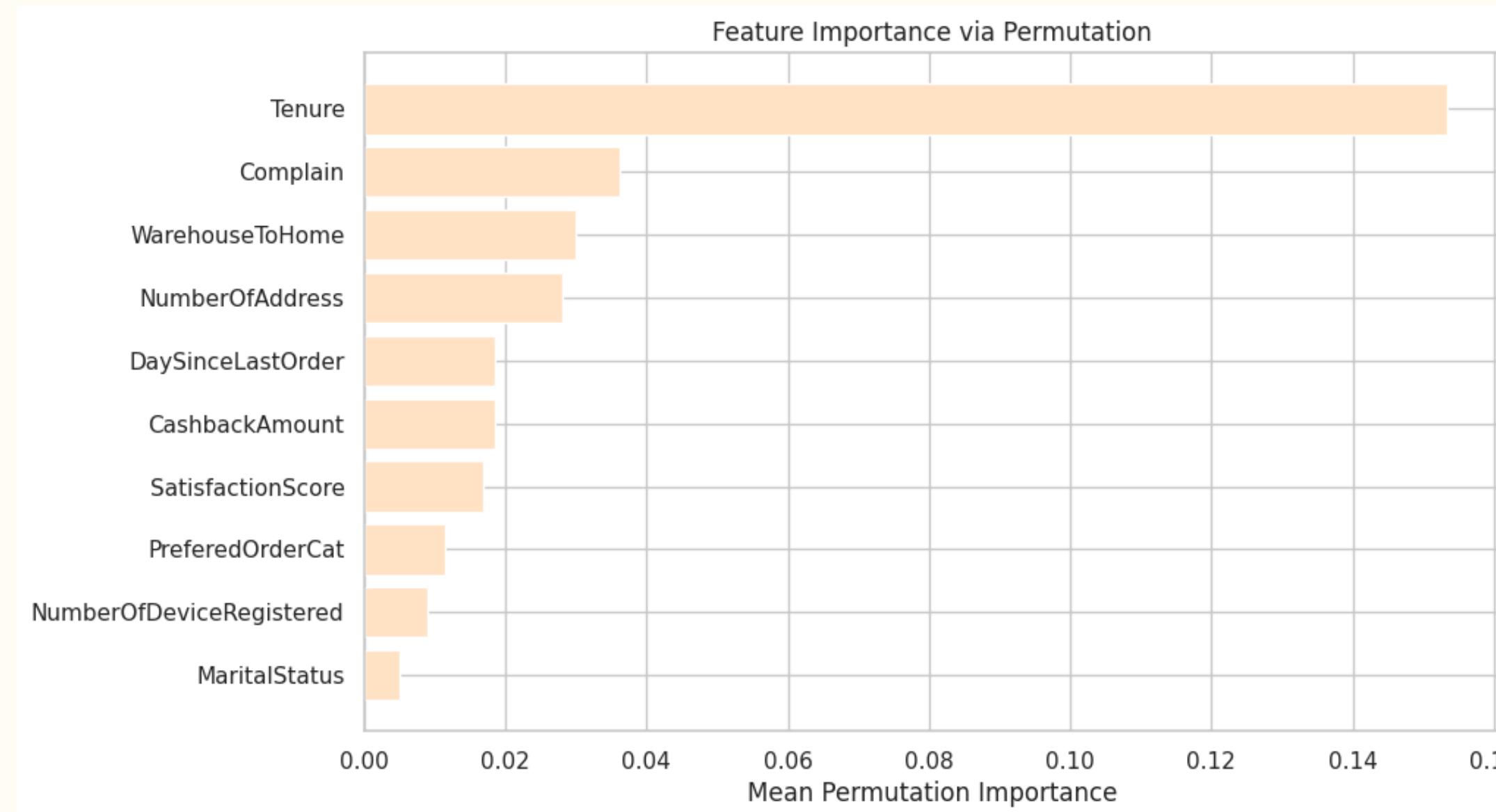
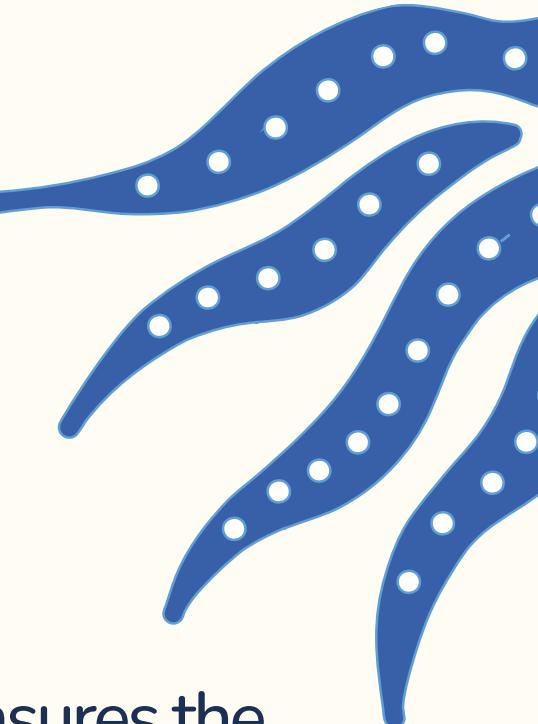
The ROC curve shows the trade-off between true positive rate (recall) and false positive rate.



Our model has an AUC of 0.90, which indicates excellent performance.



PERMUTATION IMPORTANCE



Permutation importance measures the decrease in model performance when each feature is randomly shuffled.



The most important features according to permutation importance are:

- Tenure
- Complaints
- WarehousToHome
- NumberOfAddress



MODEL LIMITATIONS

DATA LIMITATIONS

Model accuracy depends on data quality and representativeness.

FEATURE LIMITATIONS

The model may not include all possible factors influencing churn.

GENERALIZABILITY

Model performance may vary for customer segments or time periods not in the training data. Continuous monitoring and retraining are essential.

RECOMMENDATIONS

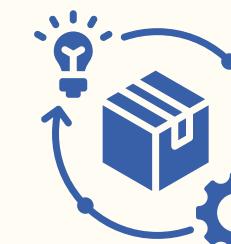
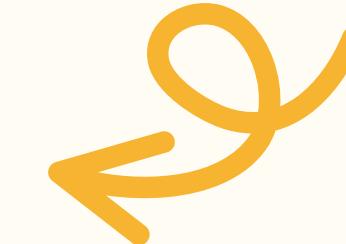


BUSINESS RECOMMENDATION



Focus On High Value Customer

Prioritize retention efforts on customers with high customer lifetime value (CLTV) to maximize returns on investment



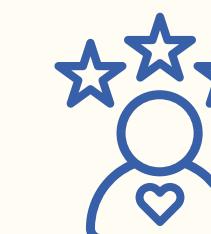
Product Improvement

Develop personalized retention campaigns for high-risk segments identified by the model
Targeted Interventions



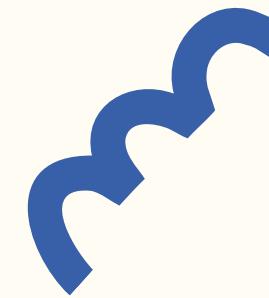
Targeted Interventions

Develop personalized retention campaigns for high-risk segments identified by the model



Customer Satisfaction

Continuously monitor and improve overall customer satisfaction



WHO AND WHEN WILL USE THIS MODEL

MAREKTING TEAM

Design targeted retention campaigns, analyze effectiveness, and optimize marketing spend

CUSTOMER SERVICE TEAM

Proactively reach out to high-risk customers identified by the model to address their concerns, resolve issues, and improve their overall experience

PRODUCT TEAM

Leverage the insights gained from the model to improve products, enhance the shopping experience, and address pain points that contribute to churn

EXECUTIVE TEAM

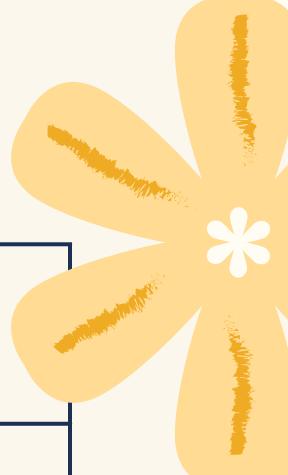
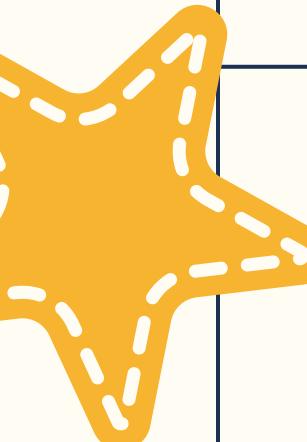
Monitor the model's impact on key business metrics like revenue, customer retention, and profitability. Use these insights to make informed, data-driven decisions regarding resource allocation and strategic planning

Regular updates to maintain accuracy.

Real-time for personalized recommendations

Periodic review to assess effectiveness and identify areas for improvement

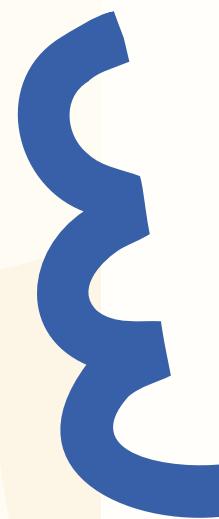
Post-purchase follow-up to encourage repeat business.



Financial Impact



- Without Model:
 - Total Loss: \$53,000
 - Total Profit: \$219,000
- With Model:
 - Total Loss: \$15,956 (-69.894%)
 - Total Profit: \$287,544 (+31.299%)



**THANK YOU
SO MUCH**

