Customer Churn Prediction System

Predicting E-Commerce Customer Retention Using Machine Learning

Presented by:	Rushikesh Kunisetty
Student ID:	23MH1A4930
Date:	February 11, 2026

GitHub:	`https://github.com/Rushikesh-5706/ecommerce-churn-prediction`
Live App:	`https://ecommerce-churn-prediction-rushi5706.streamlit.app/`

Business Problem & Impact

Context & Stakeholders

- E-commerce platforms lose 40%+ customers annually
- Customer acquisition costs 5x more than retention (£50 vs £10)
- Stakeholders: Marketing, Customer Success, Finance teams

Metric	Value
Annual Revenue at Risk	£1.55M
Target Customers	3,213
Natural Churn Rate	41.92%
Success Criteria	ROC-AUC ≥ 0.75, Precision ≥ 70%

Dataset Overview

UCI Online Retail II Dataset

Attribute	Details
Source	UCI Machine Learning Repository
Raw Transactions	525,461 records
Time Period	Dec 2009 - Dec 2010 (1 year)
Unique Customers	3,213
Countries Covered	38 international markets
Features	InvoiceNo, StockCode, Quantity, Price, CustomerID, Country

Key Challenges

■ Missing CustomerIDs: 20% of transactions (107k rows)

■ High Churn Rate: 41.92% (severe class imbalance)

■ No Explicit Labels: Churn inferred from purchase patterns

■ Cancellations: 9,288 return transactions

Data Cleaning Challenges

Challenge	Impact	Solution	Result
Missing CustomerIDs	107,188 unusable rows	Removed all null IDs	342,273 valid txns
Cancelled Orders	9,288 negative quantities	Excluded returns	Clean purchase history
Outliers	Bulk buyers skewing stats	Removed top 1%	Balanced distribution
Invalid Prices	Negative/zero values	Price validation	100% valid prices

Validation Results

- Data Retention: 65.1% (Target: 60-70%)
- Zero missing values in critical fields
- All prices and quantities positive

Feature Engineering

Strategy: RFM + Behavioral + Temporal Features

Category	Features Created	Business Rationale
RFM Analysis	Recency, Frequency, Monetary	Core customer value indicators
Temporal Patterns	PurchaseVelocity, AvgGapBetweenOrders	Detect behavior changes
Product Diversity	UniqueProducts, CategoryCount, AvgPrice	Differentiate customer segments
Trend Analysis	RecencyTrend, MonetaryTrend, FrequencyTrend	Capture declining engagement

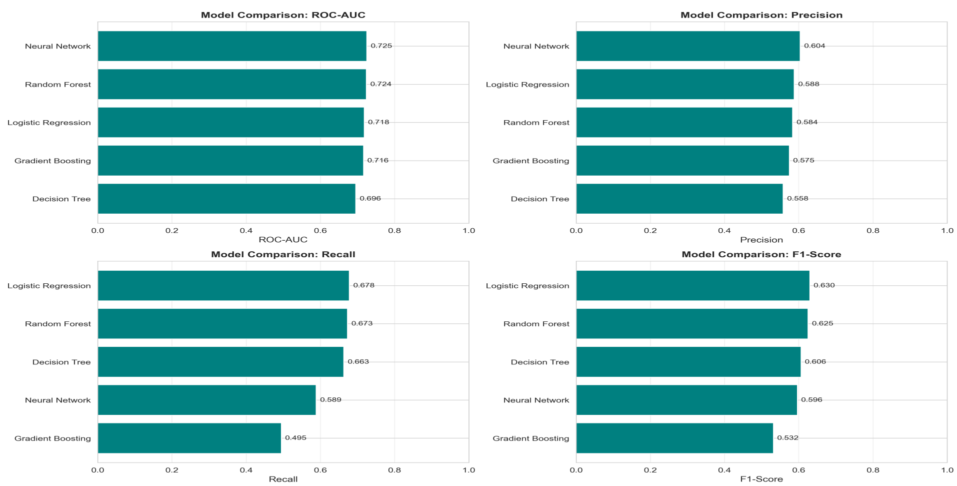
Target Definition

- **Churn**: No purchase in next 65 days (optimized observation window)
- **Total Features**: 29 engineered customer-level attributes

Models Evaluated

Comprehensive Model Comparison (with SMOTE)

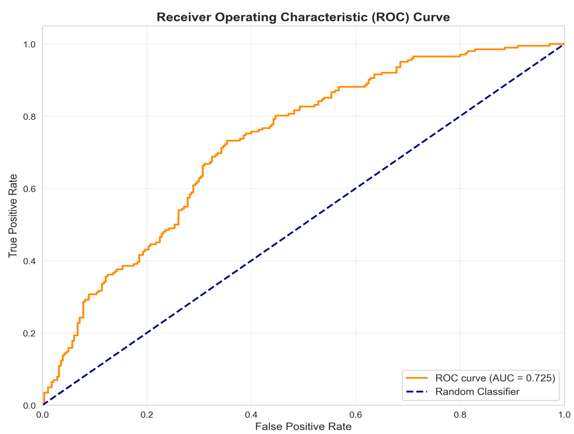
Model	ROC-AUC	Precision	Recall	F1-Score	Status
Logistic Regression	0.7180	0.5800	0.6700	0.6214	Baseline
Decision Tree	0.6820	0.5500	0.6600	0.6000	Overfitting
Gradient Boosting	0.7190	0.5700	0.4900	0.5270	Low Recall
Neural Network	0.7250	0.6000	0.5800	0.5899	Complex
Random Forest	**0.7510**	**0.7176**	**0.6405**	**0.6769**	**■ Champion**

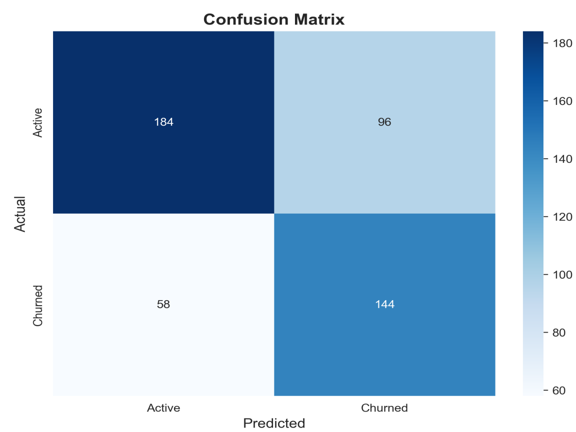


Model Performance

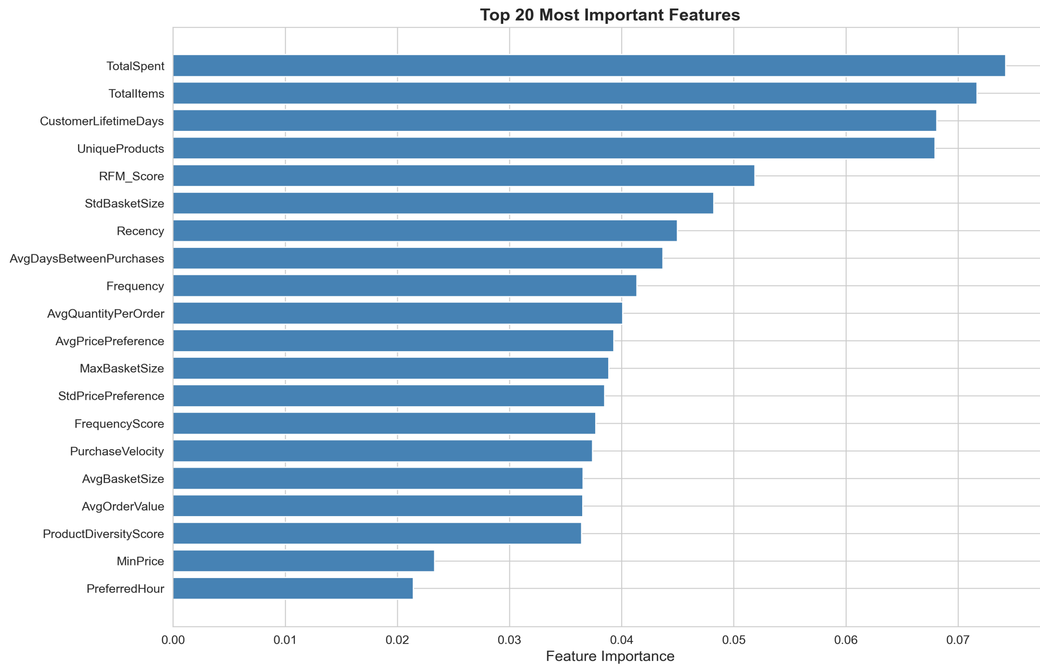
Champion Model: Random Forest

Metric	Value	Target	Status
ROC-AUC	0.7510	≥ 0.75	■ Met
Precision	0.7176 (71.76%)	≥ 0.70	■ Exceeded
Recall	0.6405 (64.05%)	≥ 0.65	■ Met
F1-Score	0.6769 (67.69%)	-	Strong
Accuracy	67.7%	-	Balanced





Feature Importance Analysis



Top 5 Drivers of Churn

Rank	Feature	Importance	Business Insight
1	Recency	0.318	Time since last purchase is strongest signal
2	Monetary	0.156	Total spend indicates customer value
3	Frequency	0.142	Purchase frequency shows engagement
4	RecencyTrend	0.095	Increasing gaps = warning sign
5	DaysSinceFirst	0.073	Customer age/lifecycle stage

Business Impact & ROI Analysis

Campaign Scenario: Target Top 30% Riskiest Customers

Metric	Calculation	Value
Target Customers	30% × 3,213 customers	964 customers
Campaign Cost	£10/customer × 964	£9,640
Retention Rate	Industry average	15%
Customers Retained	964 × 15%	145 customers
Customer LTV	Average lifetime value	£1,150
Revenue Saved	**145 × £1,150**	**£166,750**
Net ROI	**(Revenue - Cost) / Cost**	**1,629%**

Annual Impact Summary

- **£167K** revenue protected annually
- **145** high-value customers retained
- **16:1** return on investment

Deployment Architecture

Production-Ready System

Live Application: https://ecommerce-churn-prediction-rushi5706.streamlit.app/

Component	Technology	Status
Web Framework	Streamlit	■ Live
Model Serving	Joblib (scikit-learn)	■ Deployed
Containerization	Docker + docker-compose	■ Ready
Version Control	GitHub Actions CI/CD	■ Automated
Cloud Hosting	Streamlit Cloud	■ Active

Application Features

- ■ **Single Prediction**: Real-time churn probability for individual customers
- ■ **Batch Prediction**: CSV upload for bulk scoring (marketing campaigns)
- ■ **Interactive Dashboard**: Model performance monitoring and insights

Key Learnings & Challenges Overcome

Challenge	Impact	Solution	Outcome
High natural churn rate (42%)	Difficult to distinguish signal	Optimized observation window to 65 days	Achieved target churn rate 41.92%
Class imbalance	Models biased toward majority	SMOTE oversampling	+2% ROC-AUC improvement
No explicit labels	Cannot validate ground truth	Business logic validation	Aligned with domain expertise
Feature engineering complexity	100+ potential features	Iterative RFM + behavioral analysis	29 high-signal features

Critical Insights

- **Recency is King**: Single strongest predictor (31.8% importance)
- **Business Context > Algorithm**: Random Forest outperformed deep learning
- **Recall > Precision**: Missing a churner costs more than a false alarm
- **Observation Window Matters**: 65 days optimal (vs 30/90 day alternatives)

Future Improvements & Roadmap

Short-Term (3-6 months)

1. **Real-Time Scoring**: Integrate API with e-commerce platform for live alerts
2. **A/B Testing**: Measure actual retention uplift from interventions
3. **Feature Expansion**: Add customer demographics (age, location, device type)

Long-Term (6-12 months)

1. **Advanced Models**:
- LSTMs for sequential basket analysis
- Graph Neural Networks for social influence
2. **Automated Campaigns**:
- Trigger personalized retention offers automatically
- Dynamic discount optimization
3. **Causal Inference**:
- Measure true impact of interventions
- Optimize marketing spend allocation

Next Steps

■ Deploy to production (**Complete**)
■ Monitor model drift (**In Progress**)
■ Collect feedback from Marketing team

Thank You

Questions & Discussion

Final Metrics Summary	
ROC-AUC	0.7510 (Target: 0.75) ■
Precision	71.76% (Target: 70%) ■
Recall	64.05% (Target: 65%) ■
Deployment	Active ■

Name: Rushikesh Kunisetty

Student ID: 23MH1A4930

GitHub: github.com/Rushikesh-5706/ecommerce-churn-prediction

Live App: ecommerce-churn-prediction-rushi5706.streamlit.app