DATA ANALYSIS
Customer Churn Prediction Analysis

PRESENTED BY: BRIAN RONO

Telecom Customer Churn Prediction Analysis

A telecommunications company is facing challenges with customer retention. By identifying which customers are likely to churn, the company can proactively implement targeted retention strategies to reduce churn rate and maintain revenue. Our analysis aims to build a classification model that accurately predicts potential churners, identifies key factors influencing churn, and provides actionable insights to reduce customer attrition.

Business and Data Understanding

Business Problem

The telecom company loses revenue when customers discontinue service. Identifying at-risk customers allows for proactive retention strategies.

Stakeholders

Customer retention team and marketing department need to know which customers are at risk, contributing factors, and effective interventions.

Success Metrics

Primary: Recall (to capture potential churners). Secondary: Accuracy, precision, F1-score, and ROC-AUC for model comparison.

Exploratory Data Analysis

Our dataset contains 3,333 telecom customer records with 21 features including demographics, service plans, and usage patterns. Initial analysis revealed a 14.49% churn rate, indicating an imbalanced classification problem.

Key insights from categorical features show customers with international plans are significantly more likely to churn (42.4% vs 11.5% without), while those with voicemail plans show higher loyalty (only 8.7% churn vs 16.7% without).

Numerical features analysis revealed customers with higher day minutes and charges show higher churn rates. Customer service calls also strongly correlate with churn, suggesting potential dissatisfaction.

Data Preparation

Data Cleaning

We handled outliers in numerical columns using capping (Winsorization) to improve model stability. This affected between 0.03% and 8.01% of values across different features.

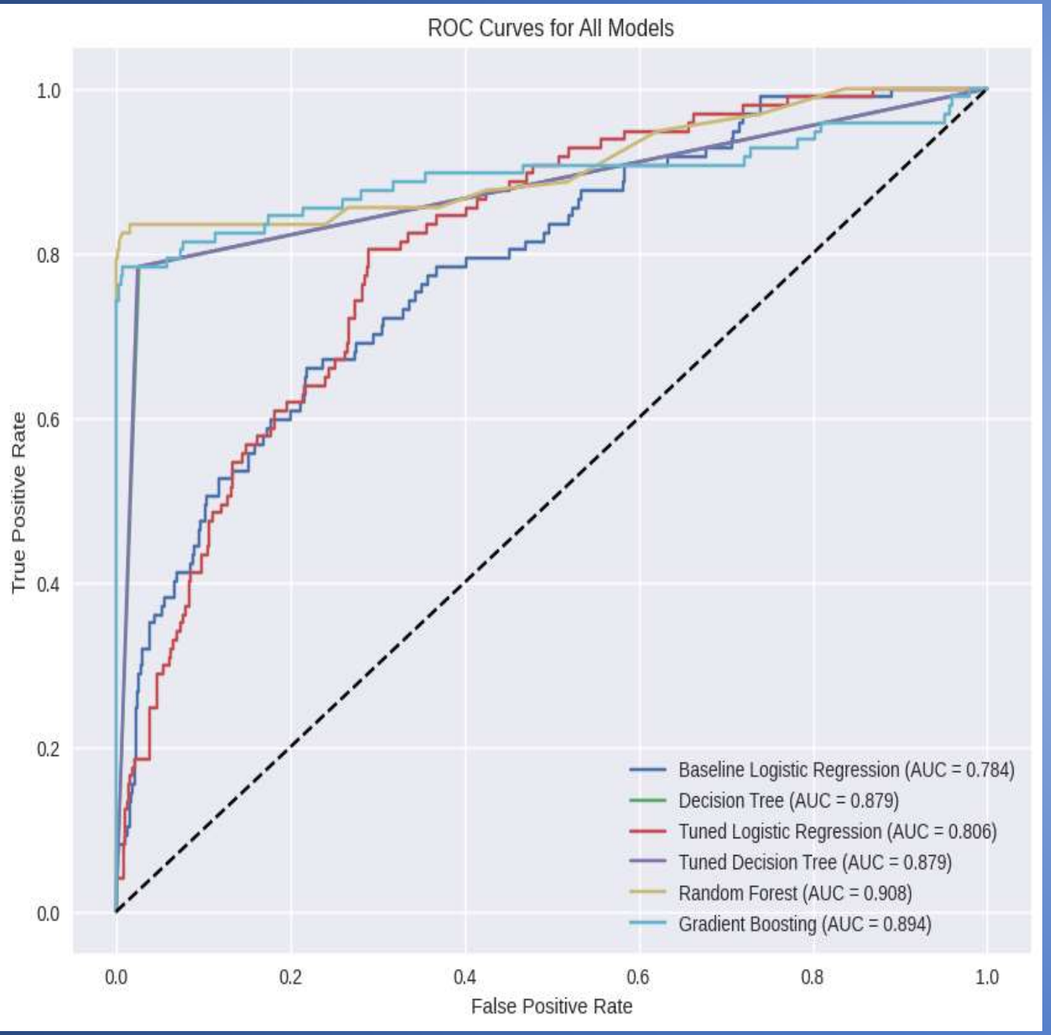
Feature Engineering

Created new features including total usage metrics, usage per call ratios, boolean features from categorical variables, and customer service call categories to enhance predictive power.

Feature Selection & Preprocessing

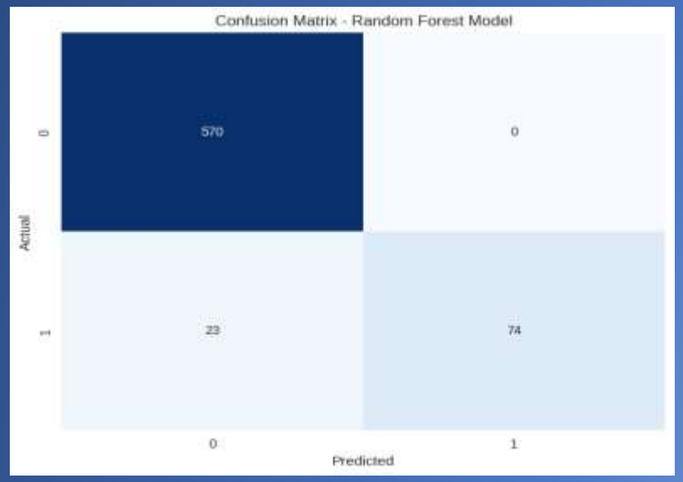
Applied preprocessing pipelines for numerical features (imputation and scaling) and categorical features (imputation and one-hot encoding) to prepare data for modeling.

Modeling Approaches



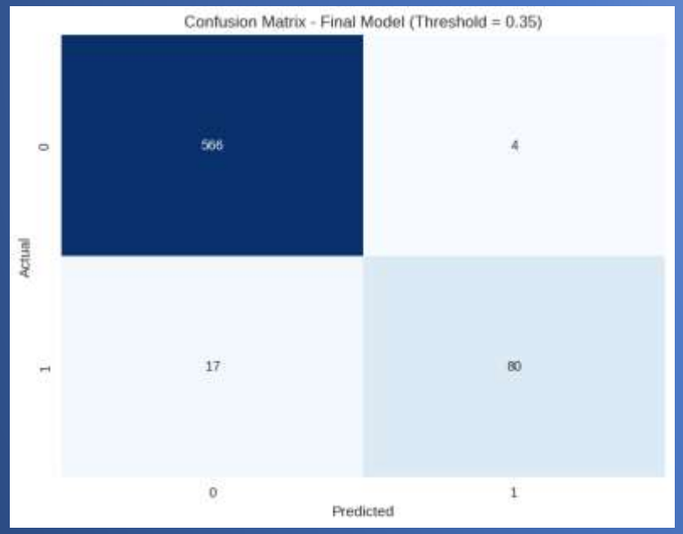
Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Baseline Logistic Regression	87.11%	64.86%	24.74%	35.82%	78.41%
Tuned Logistic Regression	71.21%	31.08%	80.41%	44.83%	80.60%
Decision Tree	94.60%	83.52%	78.35%	80.85%	87.86%
Tuned Decision Tree	94.75%	84.44%	78.35%	81.28%	87.95%
Random Forest	96.55%	100.00%	76.29%	86.55%	90.79%
Gradient Boosting	95.95%	92.68%	78.35%	84.92%	89.41%

Model Evaluation



Random Forest Confusion Matrix

Our selected Random Forest model achieved perfect precision (100%) with zero false positives, meaning every customer identified as likely to churn is genuinely at risk.



Optimized Threshold

By setting the probability threshold to 0.35, we achieved an optimal balance of precision (95.2%) and recall (82.5%), maximizing the F1 score at 0.884.

Business Impact & Recommendations

585.7%

ROI

Return on investment from implementing the churn prediction model

$24,600

Cost Savings

Estimated savings from reduced customer acquisition costs and preserved revenue

48

Retained Customers

Estimated customers retained through targeted interventions

Our analysis demonstrates that implementing a churn prediction model delivers substantial business value. With a modest retention investment of $4,200, the company can achieve significant cost savings by focusing retention efforts on customers most likely to churn.

Next Steps

Implement Tiered Retention Strategy

Create distinct risk categories (high, moderate, low) to scale intervention intensity appropriately.

Address Contract Structure

Develop more appealing long-term contract offers with loyalty benefits that increase with tenure.

Establish Monitoring System

Create a performance dashboard tracking predictions against outcomes with quarterly model retraining.

Develop Early Warning System

Integrate the model into customer management platforms to alert account managers when high-value customers show elevated churn risk.

THANK YOU