2/10/2025

Predicting Customer Churn in Telecommunications

Project Report

Khanyile, Lungisani

Table of Contents

Introduction

The telecommunications industry faces significant challenges with customer retention, as high churn rates directly impact revenue and market competitiveness. This document outlines a data-driven approach to predict customer churn and proposes actionable retention strategies. By leveraging machine learning and business intelligence tools, we aim to empower the company with insights to reduce churn and improve customer loyalty.

Problem Statement

- Current Churn Rate: 26.5% (1,869 out of 7,043 customers churned last month).
- Challenge: The company lacks a systematic method to identify customers at risk of churning, leading to reactive retention efforts.
- Impact: Unaddressed churn results in lost revenue, reduced market share, and increased customer acquisition costs.

Goals & Objectives

Goal: Develop a predictive model to identify at-risk customers and recommend retention strategies.

Objectives:

- 1. Analyze historical data to uncover patterns driving churn.
- 2. Build and evaluate machine learning models to predict churn.
- 3. Provide actionable insights to reduce churn rates.

Key Insights from Power BI Analysis

Churn Drivers

Contract & Tenure:

- Monthly contract holders churn **3x more** than annual contract customers.

- Low-tenure customers (≤6 months) are 60% more likely to churn.

Demographics:

  - Customers without dependents or partners churn: 2x more than those with family ties.
  - Senior citizens churn: 15% less than non-seniors.

Service & Payment:

  - Fibre Optic users have a **30% higher churn rate** compared to DSL users.
  - Electronic check users account for **45% of churned customers.
  - Service Gaps:
    - Customers without Tech Support, Device Protection, or Online Security are 50% more likely to churn.

Suggestions from Insights

A. Extend Contract Plans: Transition monthly contracts to 3/6-month plans.
B. Target High-Risk Segments: Focus on single customers with no dependents.
C. Enhance Service Packages: Bundle Tech Support, Device Protection, and Online Security into standard plans.

Modeling Methodology

Models Evaluated

- Logistic Regression
- Random Forest
- Gradient Boosting
- XGBoost
- Neural Network (MLPClassifier)
- Decision Tree
- AdaBoost

Optimization Techniques

- Hyper parameter Tuning: Grid Search/Random Search to optimize model parameters.
- Validation: 5-fold cross-validation to ensure robustness.

Model Performance Comparison

Model	ROC-AUC Score	Execution Time (s) |
Neural Network	0.859	336.85
Logistic Regression	0.859	32.08
XGBoost	0.857	0.81
Random Forest	0.858	739.01
Decision Tree	0.774	7.23

Key Observations

- Neural Network: Highest ROC-AUC but slowest execution.

- Logistic Regression: Optimal balance of performance (ROC-AUC: 0.859) and speed (32s).

- XGBoost: Fastest model (0.81s) with competitive ROC-AUC (0.857).

Results & Findings

Initial Models

- Top Performers:
- Neural Network (ROC-AUC: 0.859).
- Logistic Regression (ROC-AUC: 0.859, 10x faster than Neural Network).
- XGBoost (ROC-AUC: 0.857, 40x faster than Logistic Regression).

Enhanced Models

- PCA Model:
- Accuracy: 0.8069 | Execution Time: 0.66s (fastest).
- Trade-off: Lower accuracy compared to Logistic Regression.
- Regularization Model:
- Accuracy: 0.8135 | Execution Time: 17.14s.
- Best enhanced model but slower than PCA.

Recommendations

Technical Recommendations

- For High Accuracy: Deploy the Neural Network for batch predictions.
- For Real-Time Use: Prioritize Logistic Regression or XGBoost.

Business Recommendations

- Retention Campaigns: Target high-risk customers (monthly contracts, singles).
- Service Bundles: Include Tech Support and Device Protection in base plans.
- Loyalty Incentives: Offer discounts for long-term contracts.

Conclusion

The analysis highlights the critical role of predictive modelling in addressing customer churn. By combining Logistic Regression's efficiency with targeted retention strategies, the company can proactively reduce churn rates. While enhanced models (PCA, Regularization) improve speed, further hyper-parameter tuning and feature engineering could close the performance gap with initial models.

Next Steps

Model Deployment

- Integrate Logistic Regression into the customer management system for real-time alerts.

A/B Testing

- Pilot extended contracts and service bundles in high-churn regions.

Continuous Monitoring

- Track churn metrics post-intervention to refine strategies.