Proactive Customer Retention for Telecom Services

Predicting Customer Churn Using Machine Learning

Executive Summary

Customer churn is one of the most significant challenges facing telecommunications companies. Acquiring new customers costs more than retaining existing ones, making churn prediction and prevention a critical business priority.

This project developed a machine learning model to predict which customers are likely to churn, enabling proactive retention efforts. Using data from 7,043 telecommunications customers, I built and compared multiple classification models, ultimately selecting a **Gradient Boosting (tuned) classifier** that achieves **86.5% ROC-AUC** in identifying at-risk customers.

Key Finding: Customers with month-to-month contracts, tenure under 12 months, fiber optic internet service and electronic check payments are **2.7 times more likely to churn** than the average customer.

---

1. Problem Statement

Business Context

A telecommunications company is experiencing significant customer churn, resulting in lost revenue and increased customer acquisition costs. The company wants to:

1. **Identify** customers who are likely to cancel their service
2. **Understand** what factors drive customer churn
3. **Develop** targeted retention strategies for at-risk customers

Project Objective

Build a predictive model that can:

- Score customers by their likelihood to churn
- Identify the key factors that predict churn
- Enable the business to take proactive retention actions

Success Criteria

- Develop a model with ROC-AUC > 0.80 (strong discriminative ability)
- Identify actionable predictors that the business can influence
- Provide concrete recommendations for retention strategies

---

2. Data Overview

Dataset Description

The dataset contains information on 7,043 telecommunications customers, including:

- **Demographics:** Gender, senior citizen status, partner, dependents
- **Account Information:** Tenure, contract type, payment method, billing preferences
- **Services:** Phone service, internet service, streaming, security add-ons
- **Charges:** Monthly charges, total charges
- **Target Variable:** Churn (Yes/No)

Churn Distribution

Status	Count	Percentage
Retained (No Churn)	5,174	73.5%
Churned	1,869	26.5%

The dataset exhibits class imbalance, with churned customers representing approximately one-quarter of the total customer base.

---

3. Exploratory Data Analysis

Key Finding 1: Tenure is the Strongest Predictor

New customers are dramatically more likely to churn than long-term customers:

Tenure Group	Churn Rate
0-12 months	**47.4%**

Tenure Group	Churn Rate
13-24 months	28.7%
25-48 months	20.4%
49-60 months	14.4%
61+ months	**6.6%**

Insight: The first year is critical. Nearly half of customers who leave do so within their first 12 months. This represents both a challenge and an opportunity—early intervention could significantly reduce churn.

Key Finding 2: Contract Type Dramatically Affects Churn

Contract Type	Churn Rate
Month-to-month	**42.7%**
One year	11.3%
Two year	**2.8%**

Insight: Month-to-month customers churn at 15 times the rate of two-year contract customers. Contract commitment is a strong indicator of customer loyalty.

Key Finding 3: Fiber Optic Customers Have Higher Churn

Internet Service	Churn Rate
Fiber optic	**41.9%**
DSL	19.0%
No internet	7.4%

Insight: Fiber optic customers, despite having premium service, show the highest churn rate. This may indicate pricing concerns or service quality issues that warrant investigation.

Key Finding 4: Payment Method Signals Commitment

Payment Method	Churn Rate
Electronic check	**45.3%**
Mailed check	19.1%
Bank transfer (auto)	16.7%
Credit card (auto)	15.2%

Insight: Customers using electronic checks churn at nearly three times the rate of those with automatic payments. Automatic payment setup may indicate higher customer commitment and satisfaction.

High-Risk Customer Profile

Combining these factors, I identified a high-risk customer segment:

- Month-to-month contract
- Tenure ≤ 12 months
- Fiber optic internet
- Electronic check payment

This segment represents 9% of customers but accounts for 71% churn rate—2.7 times the average.

---

4. Methodology

Data Preprocessing

1. **Missing Value Treatment:** 11 records had empty TotalCharges (new customers with zero tenure). These were filled with their MonthlyCharges value.

2. **Feature Encoding:** Categorical variables were converted using one-hot encoding with `drop_first=True` to avoid multicollinearity, resulting in 30 features.

3. **Feature Scaling:** Numeric features (tenure, MonthlyCharges, TotalCharges) were standardized using StandardScaler to ensure equal weighting in distance-based algorithms.

4. **Train/Test Split:** Data was split 80/20 for training and testing, with stratification to maintain class distribution.

Model Selection

Three classification algorithms were evaluated:

Model	Rationale
K-Nearest Neighbors	Simple baseline; effective for pattern recognition
Random Forest	Ensemble method; handles mixed feature types well
Gradient Boosting	Sequential ensemble

Evaluation Metrics

Given the business context, I prioritized:

- **ROC-AUC:** Overall ability to rank customers by churn probability
- **Recall:** Ability to identify actual churners (minimize missed at-risk customers)
- **Precision:** Accuracy of churn predictions (avoid wasting resources on false alarms)

---

5. Results

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
KNN	0.766	0.579	0.488	0.529	0.778
Random Forest	0.786	0.639	0.442	0.523	0.820
Random Forest (Tuned)	0.797	0.661	0.480	0.556	0.840
Gradient Boosting	0.803	0.663	0.520	0.583	0.843

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Gradient Boosting (Tuned)	0.813	0.700	0.531	0.604	0.865

Final Model: Gradient Boosting (Tuned)

Hyperparameters:

- learning_rate: 0.05
- max_depth: 3
- n_estimators: 100
- subsample: 0.8

Performance Interpretation:

- **86.5% ROC-AUC:** Strong ability to distinguish churners from non-churners
- **70% Precision:** 7 out of 10 predicted churners will actually churn
- **53% Recall:** Model identifies about half of all actual churners
- **81% Accuracy:** Correct classification for most customers

Feature Importance

The model identified the following as the most predictive features:

Rank	Feature	Importance
1	Tenure	33.3%
2	Internet Service (Fiber optic)	20.2%
3	Payment Method (Electronic check)	9.7%
4	Contract (Two year)	8.0%
5	Contract (One year)	5.9%
6	Monthly Charges	4.8%
7	Total Charges	4.8%
8	Online Security (Yes)	2.2%

Rank	Feature	Importance
9	Paperless Billing (Yes)	2.0%
10	Streaming Movies (Yes)	1.4%

Key Insight: Tenure alone accounts for one-third of the model's predictive power, reinforcing the critical importance of early customer engagement.

---

6. Business Recommendations

Based on the analysis, I recommend the following concrete actions:

Recommendation 1: Implement a "First 90 Days" Retention Program

Rationale: With 47% of first-year customers churning, the onboarding period is critical.

Actions:

- Assign dedicated support representatives to new customers
- Schedule proactive check-in calls at 30, 60, and 90 days
- Offer loyalty incentives at the 6-month and 12-month marks
- Create educational content to help customers maximize service value

Expected Impact: Reducing first-year churn by 20% would retain approximately 415 additional customers annually.

Recommendation 2: Incentivize Contract Commitments

Rationale: Month-to-month customers churn at 15x the rate of two-year customers.

Actions:

- Offer meaningful discounts (15-25%) for one-year and two-year contracts
- Provide service upgrades or equipment credits for contract sign-ups
- Create a "contract conversion" campaign targeting long-tenured month-to-month customers
- Bundle premium features exclusively for contract customers

Expected Impact: Converting 10% of month-to-month customers to annual contracts could reduce overall churn by 3-4 percentage points.

Recommendation 3: Investigate and Address Fiber Optic Service Issues

Rationale: Fiber optic customers churn at 42% despite paying premium prices.

Actions:

- Conduct customer satisfaction surveys specifically for fiber optic users
- Analyze service quality metrics and complaint data
- Review pricing competitiveness vs. market alternatives
- Consider service level guarantees or credits for outages

Expected Impact: Identifying and fixing root causes could significantly reduce churn in this high-value customer segment.

---

7. Ideas for Further Research

1. **Survival Analysis:** Model not just whether customers will churn, but when, to optimize intervention timing.

2. **Customer Lifetime Value Integration:** Combine churn probability with customer value to prioritize retention efforts on high-value at-risk customers.

3. **Natural Language Processing:** Analyze customer service call transcripts and chat logs to identify early warning signals of dissatisfaction.

4. **A/B Testing Framework:** Design experiments to measure the effectiveness of different retention interventions on high-risk segments.

5. **Segment-Specific Models:** Build separate models for different customer segments (e.g., senior citizens, business customers) that may have different churn patterns.

---

8. Conclusion

This project successfully developed a machine learning model to predict customer churn with 86.5% ROC-AUC, exceeding the 80% target. The analysis revealed that:

- **Tenure is paramount:** The first year is make-or-break for customer retention
- **Contracts drive loyalty:** Longer commitments correlate with dramatically lower churn
- **Service choices matter:** Fiber optic and electronic check payment are risk indicators

By implementing the recommended retention strategies—focusing on new customer onboarding, incentivizing contracts, and investigating fiber optic service issues—the company can proactively reduce churn and improve customer lifetime value.