

■ Customer Churn Prediction Case Study

1. Project Overview

Customer churn is a critical issue for telecommunications companies. Losing customers directly reduces revenue and increases acquisition costs. This project analyzes the Telco Customer Churn dataset (7,043 customers, 21 variables) to predict churn likelihood, identify the key factors driving customer attrition, and recommend strategies to improve retention.

2. Dataset

Source: IBM Sample Data / Kaggle (Telco Customer Churn). Size: 7,043 rows × 21 columns.

Key Features:

- Customer demographics (gender, senior citizen, dependents)
- Account information (tenure, contract type, payment method)
- Services (internet, phone, streaming)
- Charges (MonthlyCharges, TotalCharges)

Target Variable: Churn (Yes/No)

3. Methodology

3.1 Data Preparation

- Cleaned missing values in TotalCharges.
- Converted categorical variables (Contract, PaymentMethod) to dummy variables.
- Balanced dataset with churn ≈ 26%.

3.2 Exploratory Data Analysis

- Tenure: Customers with <12 months tenure have highest churn.
- Contract: Month-to-month contracts show ~43% churn vs ~11% for 2-year contracts.
- Charges: High monthly charges strongly correlate with churn.

3.3 Modeling

Two models were tested:

- CART (Classification and Regression Tree): Accuracy ~82%. Clear decision rules.
- Logistic Regression: Accuracy ~80%. Confirmed predictor significance.

4. Results

Top Churn Predictors:

1. Contract type (month-to-month is riskiest)
2. Monthly charges (higher bills → higher churn)
3. Tenure (shorter tenure → higher churn)

Model Comparison:

- CART slightly outperformed logistic regression.

- CART was easier to explain to stakeholders.

5. Business Recommendations

- Promote Long-Term Contracts: Offer discounts for 1–2 year contracts.
- Customer Retention Campaigns: Target month-to-month customers with <1 year tenure.
- Billing Adjustments: Loyalty discounts or bundles for high-charge customers.
- Proactive Service: Support for early-tenure customers.

6. Limitations & Future Work

Dataset lacks qualitative data (customer satisfaction, reasons for churn).

Could integrate time-series retention tracking.

Future models: Random Forest / Gradient Boosting.

7. Conclusion

Data-driven insights can guide retention strategy. Identifying at-risk customers early reduces churn and increases profitability.

Key Takeaway: Contract type and tenure are the strongest levers for churn management.