Telecom Customer Churn Prediction

Customer churn is a critical issue in the telecom industry, directly impacting revenue, customer acquisition costs, and profitability. Identifying at-risk customers and implementing retention strategies is key to sustaining business growth.

This presentation explores our approach to predicting customer churn, understanding its key drivers, and developing actionable retention strategies based on data-driven insights.

 **Presented by Joan Gathoni**

Business Understanding & Problem Statement

The Challenge

The company is losing customers and needs a way to predict churn before it happens. Understanding why customers leave will help in making data-driven decisions to improve retention.

Our Approach

Classification task: Predict whether a customer will churn based on service usage, customer service interactions, and other factors.

Business Impact

By identifying high-risk customers early, we can implement targeted retention strategies, improve customer service, and enhance loyalty through personalized offers.

Research Questions & Objectives

1

Identify Churn Drivers

What are the most significant factors influencing customer churn? Do customers with an international plan have a higher churn rate?

2

Analyze Service Interactions

Does frequent customer service interaction indicate a higher risk of churn? How do service plans impact customer satisfaction?

3

Develop Predictive Model

Can a machine learning model accurately predict churn using available features? Which model performs best for our specific dataset?

4

Create Actionable Strategies

Based on model insights, what retention strategies should we implement to reduce churn and improve customer loyalty?

Data Overview & Preparation

Dataset Characteristics

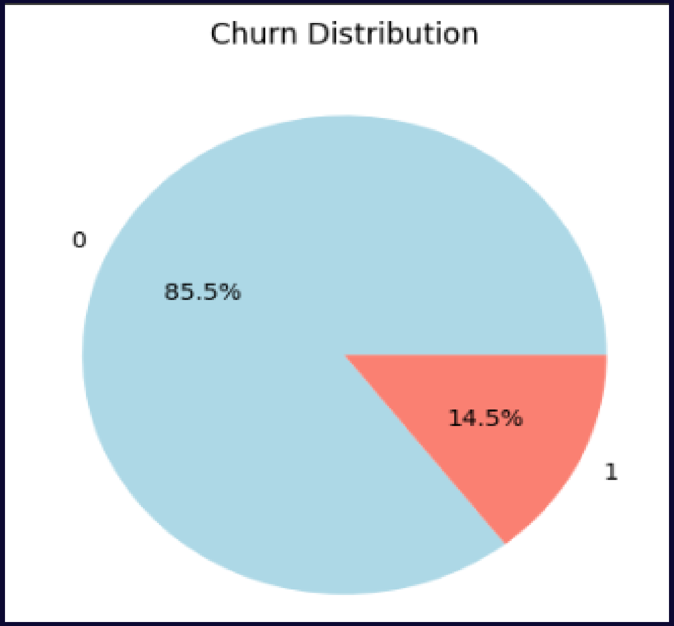
3,333 customer records with 21 features including account information, service usage patterns, and churn status. Key features include call minutes, service calls, and plan types.

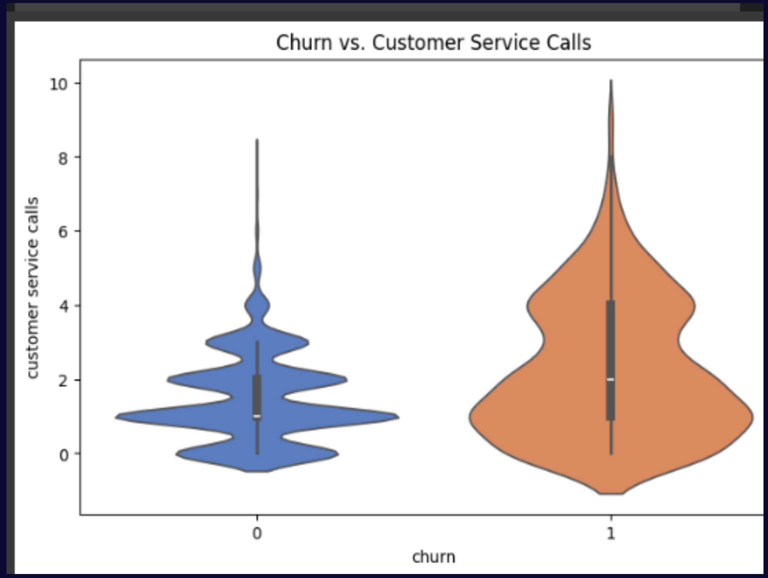
Data quality was excellent with no missing values or duplicates, though some features required transformation and encoding.

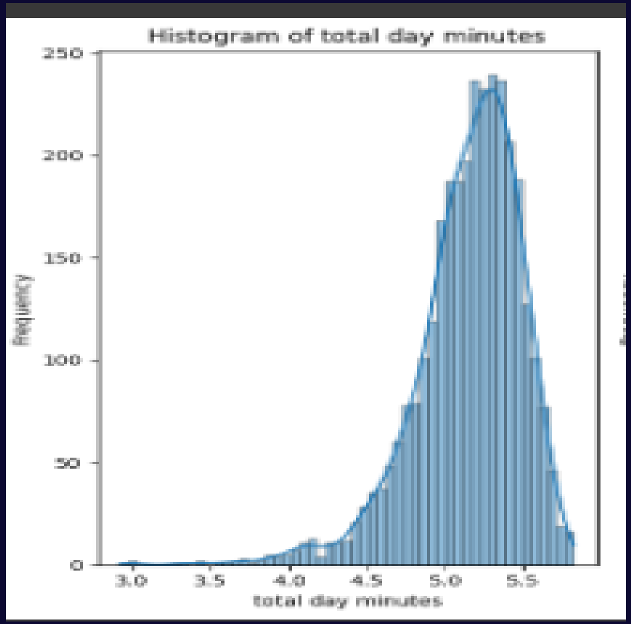
Data Preparation Steps

- Converted categorical variables (plans, churn) to numeric
- Applied one-hot encoding to state variables
- Removed irrelevant features (phone number, area code)
- Handled outliers using z-score filtering
- Applied log transformation to skewed features

Exploratory Data Analysis







Churn Distribution

The dataset shows class imbalance with approximately 15% of customers churning. This imbalance requires careful model selection and evaluation metrics.

Service Calls & Churn

Customers with more than 3 service calls show significantly higher churn rates, indicating service dissatisfaction as a key churn driver.

Usaąe Patterns

High-usage customers (total day minutes) show elevated churn rates, suggesting potential pricing dissatisfaction among heavy users.

Feature Engineering & Selection

Feature Creation

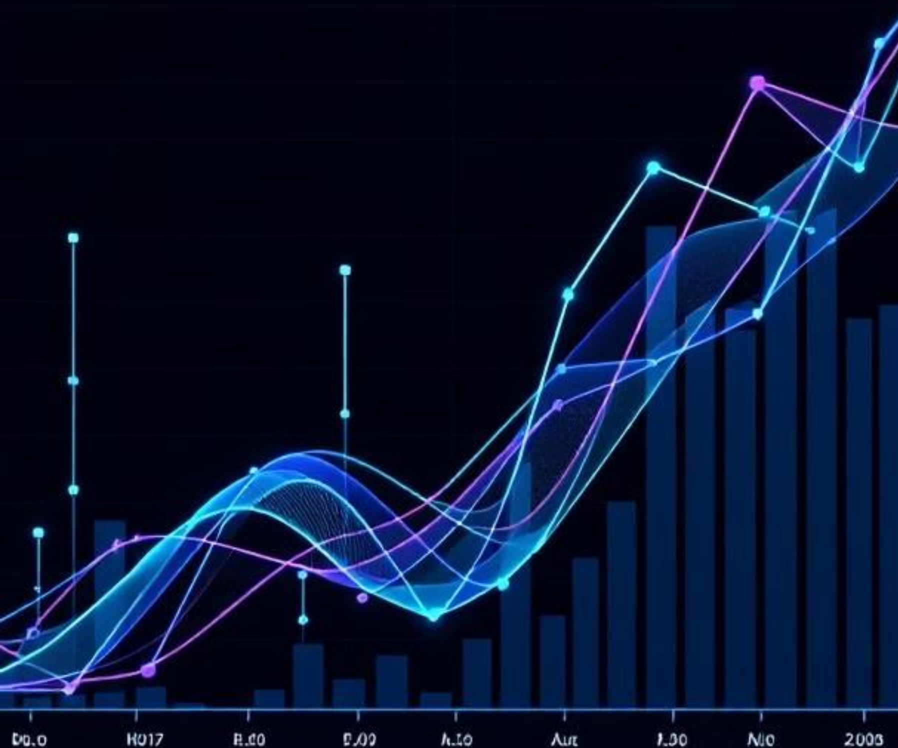
Created new features to improve predictive power, including call ratios (calls per minute), total usage minutes across all periods, and a high service calls flag for customers with more than 3 service interactions.

Correlation Analysis

Identified and removed highly correlated features (correlation > 0.9) to reduce multicollinearity and improve model efficiency. Features like voicemail messages and day charges were removed due to redundancy.

Feature Scaling

Applied MinMaxScaler to normalize features between 0-1 for distance-based algorithms like K-NN, and StandardScaler for other algorithms to ensure consistent feature scales.



Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	86.5%	0%	0%	0%
Decision Tree	92.0%	79%	69%	74%
Random Forest	93.0%	91%	59%	72%
K-Nearest Neighbors	84.1%	0%	0%	0%

The Decision Tree model offers the best balance between precision and recall with an F1 score of 74%, while Random Forest achieves the highest overall accuracy at 93% and precision at 91%.

Key Findings & Recommendations

1 Primary Churn Drivers

Customer service calls (>3) strongly predict churn, indicating service dissatisfaction. High usage customers and those with international plans also show elevated churn rates. States like NJ, CA, and TX have significantly higher churn rates.

2 Recommended Retention Strategies

Implement proactive outreach after second service call to address issues before they escalate. Review pricing for high-usage customers and international plans. Develop targeted retention offers for high-risk segments and high-churn states.

3 Implementation Plan

Deploy the Decision Tree model for operational use due to its balanced performance. Create an automated alert system for customers crossing risk thresholds. Establish a feedback loop to continuously improve model accuracy and retention effectiveness.

Contact Info:
LinkedIn: www.linkedin.com/in/joan-b1aa8023b
Mobile: 0725145984
Email; gathonigatharia@gmail.com