

---

# Comprehensive Analysis and Predictive Modeling of Customer Churn

Prepared by: Dennis Muuo

## Overview

This report covers the analysis of customer churn data from a telecommunications company. It includes an exploratory data analysis (EDA) to identify patterns in customer behavior, followed by the development of a predictive model to understand the factors contributing to churn. The goal is to build an effective model for predicting which customers are likely to churn, enabling the company to implement proactive retention strategies.

## 1. Business Problem

### 1.1 Business Context

Customer churn is a significant issue for telecommunications companies. When customers stop using the company's services, it leads to loss of revenue and a negative impact on business growth. Reducing churn is critical to ensuring long-term profitability and customer loyalty. This project aims to build a model that predicts customer churn based on various features such as usage statistics and interactions with the company.

### 1.2 Business Question

- How can we predict which customers are likely to churn based on their usage data and interactions?
- Which features most significantly contribute to the likelihood of churn?

### 1.3 Stakeholder Impact

- **Why is this important?** Accurate predictions of churn allow the company to implement preemptive retention strategies and targeted communication efforts.
- **Use Cases:**
  - Personalized offers and campaigns to retain customers.
  - Optimizing customer service operations by identifying high-risk customers early.

## 2. Exploratory Data Analysis (EDA)

### 2.1 Data Understanding

The dataset used in this analysis contains the following features:

- **Customer Details:** Includes demographic and personal information such as state, account length, and area code.
- **Account Features:** Information about subscription plans and customer service interactions.
- **Usage Statistics:** Data on call usage during different times of the day (day, evening, night, international).
- **Churn Status:** The target variable indicating whether a customer has churned.

## 2.2 Data Cleaning

Upon reviewing the dataset, there were no missing values or duplicates, ensuring a clean dataset for analysis.

## 2.3 Key Findings from EDA:

- **Distribution of Churn:** The churn rate was found to be relatively low.
- **Churn and Account Length:** Churned customers tended to have a longer account length.
- **Churn and Usage Patterns:** Churned customers used more call minutes (day, evening, and night).
- **Impact of Plans on Churn:** Customers with international calling plans had higher churn rates, while those with voicemail plans had lower churn rates.
- **Customer Service Calls:** Customers who frequently contacted customer service were more likely to churn.

## 2.4 Insights and Recommendations:

- **Pricing Strategy:** Consider revising pricing models, especially for high-usage customers.
- **International Call Services:** Enhance international calling offerings to reduce churn among this group.
- **Retention Programs for Long-Term Customers:** Focus retention efforts on long-term customers.
- **Promote Voicemail Plans:** Leverage the positive impact of voicemail plans on retention.

# 3. Predictive Modeling

## 3.1 Feature Selection

The following features were selected for model training:

- `total day minutes, total eve minutes, total night minutes, total intl minutes, customer service calls, and number vmail messages.`

Redundant features such as `total day charge` and `total night charge` were excluded to avoid multicollinearity.

## 3.2 Model Development

A logistic regression model was built to predict churn, and hyperparameter tuning was conducted using GridSearchCV to enhance model performance.

#### *Baseline Model:*

- **Accuracy:** 0.82
- **ROC AUC Score:** 0.72

#### *Tuned Model:*

- **Accuracy:** 0.83
- **ROC AUC Score:** 0.74

### **3.3 Model Evaluation:**

The tuned logistic regression model provided better performance than the baseline model, indicating improved predictive power after hyperparameter optimization.

### **3.4 Key Features Influencing Churn:**

- **Total Day Minutes:** Higher usage of day minutes was positively correlated with churn.
- **Customer Service Calls:** Frequent interactions with customer service increased churn risk.
- **International Call Usage:** Customers using international calls were more likely to churn.

### **3.5 Business Implications:**

- **Customer Segmentation:** Use the model to identify at-risk customers and implement targeted retention strategies.
- **Pricing and Plans:** Focus on customers with high usage and international call needs.

## **4. Conclusion and Recommendations**

### **4.1 Key Findings:**

- Churn is influenced by factors such as high call usage, frequent customer service calls, and international call plans.
- The logistic regression model provides valuable insights into the key drivers of churn, with a reasonable level of accuracy in predicting customer behavior.

### **4.2 Recommendations:**

- **Loyalty Programs:** Implement loyalty programs to reward long-term customers and mitigate churn.
- **Improved Customer Support:** Enhance the customer service experience to reduce churn due to dissatisfaction.
- **Targeted Retention Efforts:** Use predictive models to proactively identify at-risk customers and offer personalized retention strategies.

**Thankyou**