SimuTel Churn Prediction

Data-Driven Insights for Customer Retention

BY: MAGGIE KURIA
TITLE: DATA SCIENTIST

Table of contents

01

INTRODUCTION

Business Problem

- ❑ **Challenge:** SimuTel is experiencing customer churn but lacks tools to identify at-risk customers in advance.
- ❑ **Impact:** Losing customers increases acquisition costs and reduces profitability
- ❑ **Goal:** Build a predictive model to flag customers likely to churn and guide retention actions.

Project Objectives

- Clean and prepare telecom customer data
- Understand churn drivers using historical features
- Understand churn drivers using customer data
- Build and evaluate classification models
- Generate churn risk scores
- Recommend actions to reduce churn

02

DATA EXPLORATION

Dataset Overview

Data Source

- Kaggle Telecom Churn Dataset
- 3,333 customer records
- Target Variable: Churn (Yes/No)

Features

- Plan types (international, voicemail)
- Call & charge metrics
- Customer service calls
- State & area code

Data Cleaning

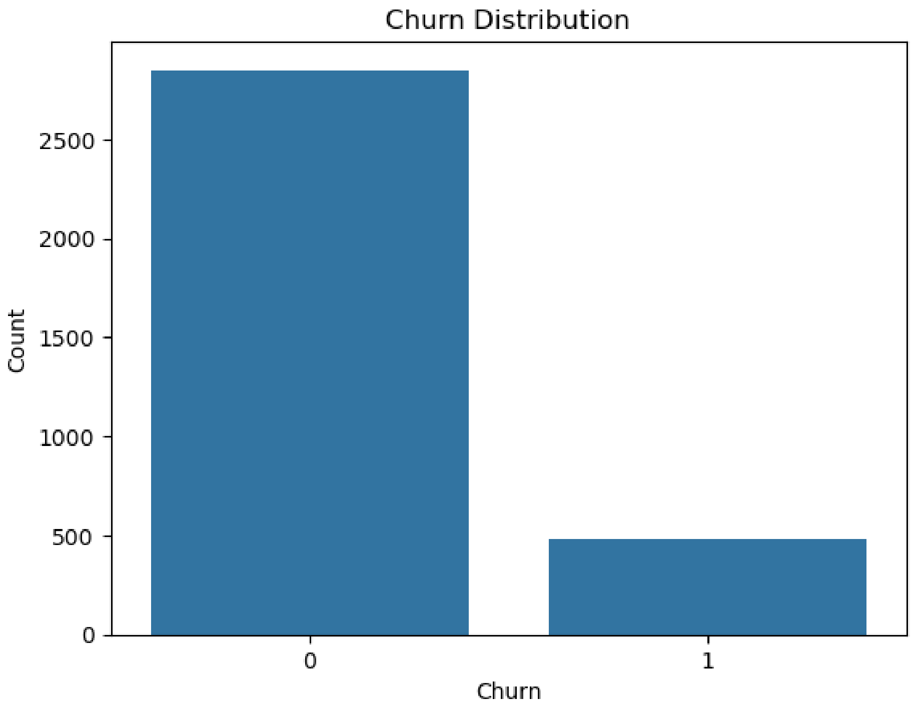
- Dropped irrelevant fields (e.g., phone number)
- Converted 'yes/no' fields to binary (1/0)
- One-hot encoded state and area code
- Scaled features for linear models (e.g., Logistic Regression)

Exploratory Analysis

- Majority of customers have not churned, showing a class imbalance.

- Only 14.5% of customers churned

- Strong churn predictors based on correlation heatmap:
  - Having an international plan
  - High total day charge
  - Frequent customer service calls



03

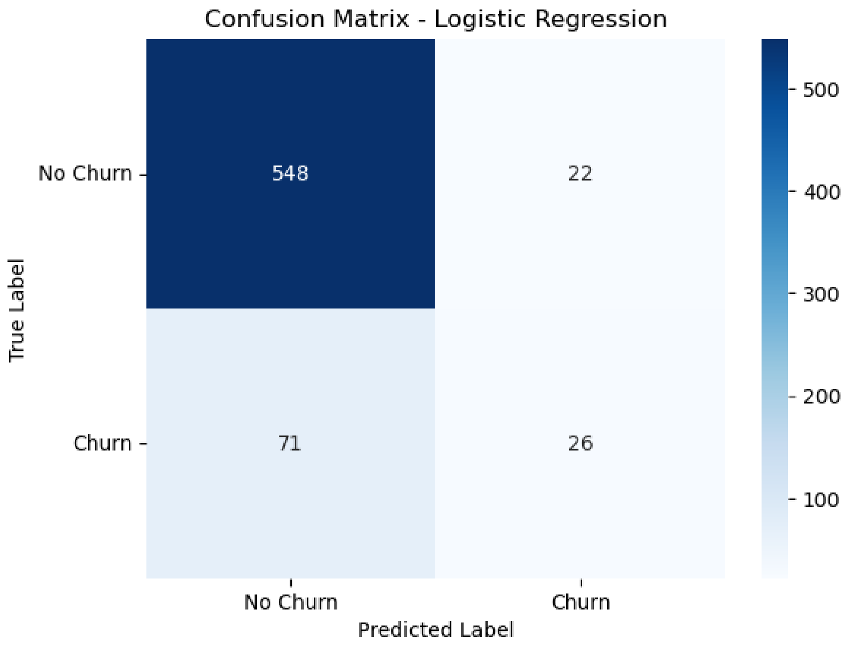
CLASSIFICATION MODELS

Model Approach

- **Baseline Model:** Logistic Regression for interpretability
- **Advanced Model:** Random Forest for performance and feature analysis
- **Evaluation Metrics:**
  - Precision
  - Recall
  - F1-score
  - Confusion Matrix
  - ROC AUC Score

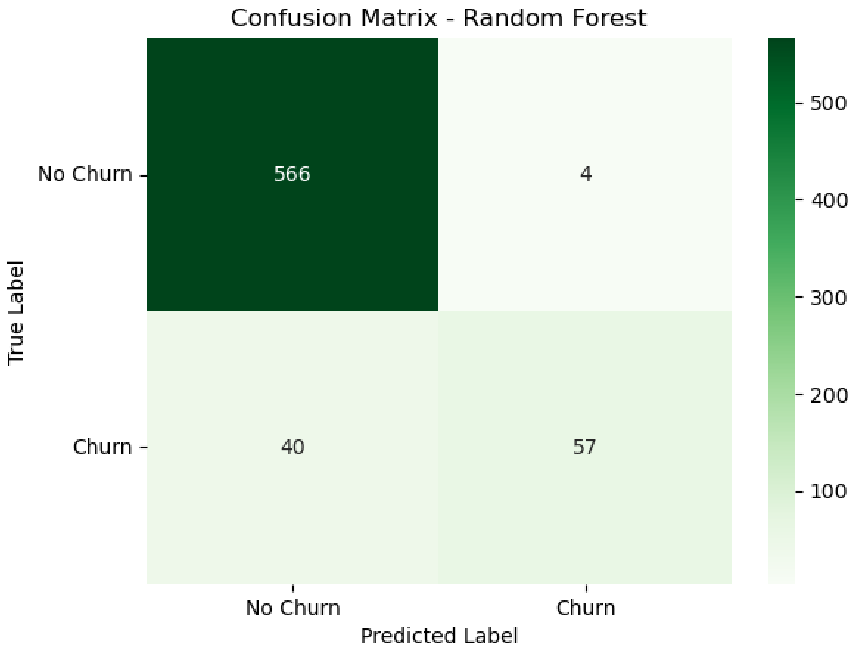
Logistic Regression

- **Precision:** 56% of predicted churners were actually churners.
- **Recall:** 0.31 Missed ~69% of actual churners.
- **F1-score:** 0.40 Balance between catching and correctly identifying churners.
- **ROC AUC:** 0.85 Strong ranking ability but misses many churners.



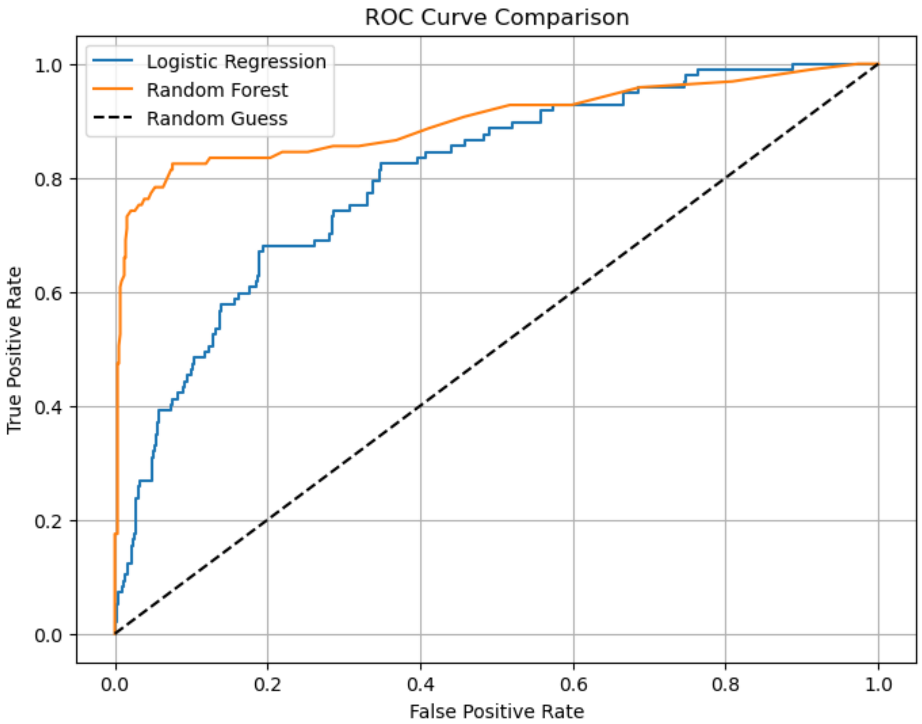
Random Forest

- **Precision:** 0.61 More accurate churn predictions.
- **Recall:** 0.51 Captures over half of all actual churners.
- **F1-score:** 0.56 Higher overall effectiveness in detecting churn.
- **ROC AUC:** 0.94 Excellent separation between churners and non-churners.

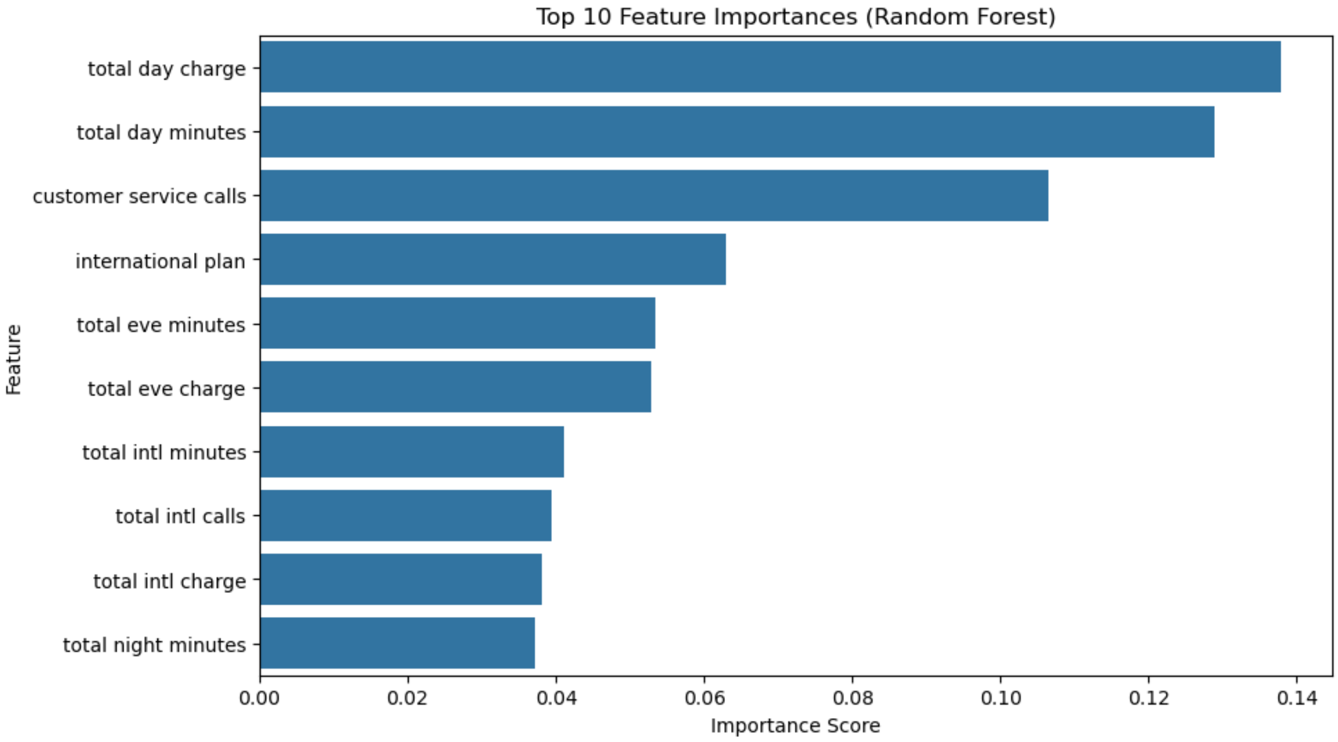


Model Performance

- Logistic Regression offers simplicity and interpretability.
- Random Forest is significantly better at identifying churners and minimizing business risk.



Feature Importance



04

CONCLUSION

Business Recommendations

- **Proactively engage customers with:**
  - International plans
  - High daytime usage
  - Frequent support calls

- **Use model scores to:**
  - Prioritize outreach
  - Personalize retention offers

- **Choose model based on:**
  - Performance: Random Forest
  - Transparency: Logistic Regression

Next Steps

- ✓ Test model on live or recent customer data
- ✓ Automate weekly churn risk scoring
- ✓ Integrate into CRM marketing platforms
- ✓ Track success of retention interventions

Thanks!

Do you have any questions?
maggiekuria72@gmail.com

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**