

Report : Predictive Modeling for Customer Churn Analysis

Date: 2nd September, 2023

Prepared by: Aishwarya Jethani

Executive Summary

This comprehensive report presents the findings and recommendations from a predictive modelling project aimed at understanding and reducing customer churn for a telecom company. The analysis delves into the telecom dataset, encompassing customer profiles, usage patterns, and churn behaviour. The ultimate goal is to provide actionable insights to minimize customer churn and enhance business profitability.

1. Introduction

1.1 Background

The telecom industry is characterized by high customer turnover, making customer churn a critical issue. A proactive approach to identify and address churn factors can significantly impact a company's bottom line. This analysis endeavours to uncover the underlying causes of customer churn and devise strategies to mitigate it.

1.2 Objectives

The analysis is driven by the following objectives:

- **Identify significant features influencing customer churn.**
- **Develop predictive models to identify customers at high risk of churning.**
- **Provide actionable recommendations for reducing churn based on data-driven insights.**

2. Data Exploration and Preprocessing

2.1 Data Loading

The analysis commenced with the ingestion of the telecom dataset, housing a wealth of information about customer profiles, usage patterns, and churn events.

2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis revealed vital insights, including:

- **A dataset containing 667 customer profiles.**
- **Distribution of the target variable, "Churn."**

2.3 Data Preprocessing

Data preprocessing steps were executed to ensure the integrity of the analysis:

- **Addressed missing values:** Careful handling of missing data to prevent adverse effects on model performance.
- **Outlier Detection and Treatment:** Employed the Interquartile Range (IQR) method to detect and manage outliers.

- **Visualization of the Target Variable (Churn):** Visualized the distribution of the target variable to understand its prevalence in the dataset.

2.4 Feature Engineering

- **Categorical Variable Encoding:** Transformed categorical variables, such as "International plan" and "Voice mail plan," into numerical format using one-hot encoding.
- **Feature Scaling:** Applied feature scaling where necessary to standardize feature scales.

3. Model Building

3.1 Splitting Data

The dataset was divided into training and testing sets to facilitate model development and evaluation.

3.2 Machine Learning Models

Implemented an array of machine learning algorithms to assess their performance:

- **Decision Tree Classifier**
- **K Nearest Neighbors (KNN)**
- **Neural Network**
- **Random Forest Classifier**

3.3 Hyperparameter Tuning

Tuned model hyperparameters using techniques such as grid search and random search to optimize model performance.

4. Model Evaluation

4.1 Performance Metrics

Decision Tree Classifier (Best Performing Model)

- Accuracy: 95.78%
- Precision: 81.58%
- Recall: 72.09%
- F1-score: 95.66%
- ROC AUC: 0.8519

K Nearest Neighbors (KNN)

- Accuracy: 90.44%
- Precision: 50.00%
- Recall: 9.30%
- F1-score: 87.36%
- ROC AUC: 0.5416

Neural Network

- Accuracy: 90.44%
- Precision: 0.00%
- Recall: 0.00%
- F1-score: 85.91%

- ROC AUC: 0.8272

Random Forest Classifier

- Accuracy: 96.89%
- Precision: 93.94%
- Recall: 72.09%
- F1-score: 96.70%
- ROC AUC: 0.8580

Transfer Learning Model (Bonus Task)

- Accuracy: 90.44%
- Precision: 0.00%
- Recall: 0.00%
- F1-score: 85.91%
- ROC AUC: Not Applicable (Non-image data)

4.1.1 Model Performance Summary

Detailed summary of model performance:

- **Decision Tree Classifier:** Delivered strong accuracy (95.78%), precision (81.58%), and recall (72.09%).
- **K Nearest Neighbors (KNN):** Faced challenges due to class imbalance, resulting in limited performance.
- **Neural Network:** Displayed low precision and recall.
- **Random Forest Classifier:** Emerged as the best-performing model with high accuracy (96.89%) and F1-score (96.70%).

4.2 ROC Curve Analysis

Plotted ROC curves and calculated the AUC for each model to assess their discrimination ability. The ROC AUC values provide insights into the models' ability to distinguish between churn and non-churn customers.

4.3 Confusion Matrices

Visualization of model performance through confusion matrices, providing insights into true positives, true negatives, false positives, and false negatives.

5. Feature Importance and Interpretation

5.1 Top Features Indicative of Churn

- **Total day minutes (Importance: 0.1414):** Higher total day usage minutes indicate a higher churn risk, suggesting customers who extensively use daytime telecom services are more likely to churn.
- **Total day charge (Importance: 0.1265):** Elevated daytime charges correlate with churn, implying that customers with higher daytime costs may consider switching to alternative providers.

- **International plan (Importance: 0.0834):** The presence of international plans increases churn likelihood, implying that the international plan might not meet customer expectations or needs effectively.

5.2 Insights into Feature Significance

These findings underscore that customers who frequently make service calls, have higher total day usage, and those with international plans are more prone to churn. This insight forms the basis for retention strategies.

6. Recommendations to Reduce Customer Churn

6.1 Actionable Strategies

Optimize Daytime Usage Plans:

- *Insights:* Total day minutes and total day charges strongly predict churn, with higher daytime usage customers at greater risk.
- *Recommendation:* Tailor daytime usage plans to cater to distinct customer segments, offering competitive pricing for high daytime usage to incentivize retention.

Review International Plans:

- *Insights:* The presence of international plans significantly influences churn.
- *Recommendation:* Conduct a comprehensive review of international plans to ensure they align with customer expectations and offer competitive rates. Solicit customer feedback to refine international offerings.

Implement Retention Offers:

- *Insights:* Customers demonstrating churn-related behaviors, such as frequent customer service calls or increased daytime usage, may benefit from retention incentives.
- *Recommendation:* Identify at-risk customers and extend retention incentives, which may include discounts on international plans, contract extensions with special offers, or loyalty rewards. Proactive engagement can enhance customer loyalty and reduce churn.

7. Conclusion

In closing, this extensive analysis provides a deep understanding of customer churn drivers and offers actionable strategies for churn reduction. The Random Forest Classifier emerged as the best-performing model. By implementing the recommended strategies, the telecom company can proactively retain customers, elevate satisfaction levels, and ultimately drive business growth.