Enhancing Customer Loyalty at Syriatel: Predictive Modeling for Churn Prevention

Business Understanding

The Main objective for this notebook is to develop an accurate predictive classifier to identify customer churn for Syriatel, a telecommunications company. The key focus is to assist SyriaTel in mitigating revenues losses by gaining insights into the factors driving customer churn.

Data processing was conducted to rectify missing values, encode categorical variables, and standardize numerical attributes, ensuring the dataset was prepared for modeling. I conducted an evaluation of four different classification models to assess their performance. After this evaluation, I selected the best-performing model among them. Following this, I examined the most important features in the chosen model.

SyriaTel faces a pressing challenge in customer churn, which can lead to substantial financial losses. Customer retention is a top priority, and this project's outcome will empower SyriaTel to proactively address churn, boost customer satisfaction, and maintain its competitive edge in the telecommunications industry.

Data Understanding

The dataset was sourced from Kaggle and includes valuable insights into customer behaviour patterns. It offers a comprehensive view of features that will be used to predict and understand some of the factors that may lead to the termination of their contract with SyriaTel.The data has exactly 3,333 rows and 21 columns The feature columns used for the analysis and predictive modeling include :state, area code, international plan, voice mail plan, number vmail messages, total day minutes, total day calls, total day charge, total eve minutes, total eve calls, total eve charge, total night minutes, total night calls, total night charge, total intl minutes, total intl calls.

Below is a brief description of the features:

state: the state the customer lives in

account length: the number of days the customer has had an account

area code: the area code of the customer

phone number: the phone number of the customer

international plan: true if the customer has the international plan, otherwise false

voice mail plan: true if the customer has the voice mail plan, otherwise false

number vmail messages: the number of voicemails the customer has sent

total day minutes: total number of minutes the customer has been in calls during the day

total day calls: total number of calls the user has done during the day

total day charge: total amount of money the customer was charged by the Telecom company for calls during the day

total eve minutes: total number of minutes the customer has been in calls during the evening

total eve calls: total number of calls the customer has done during the evening

total eve charge: total amount of money the customer was charged by the Telecom company for calls during the evening

total night minutes: total number of minutes the customer has been in calls during the night

total night calls: total number of calls the customer has done during the night

total night charge: total amount of money the customer was charged by the Telecom company for calls during the night

total intl minutes: total number of minutes the user has been in international calls

total intl calls: total number of international calls the customer has done

total intl charge: total amount of money the customer was charged by the Telecom company for international calls

customer service calls: number of calls the customer has made to customer service

churn: true if the customer terminated their contract, otherwise false

The data underwent data processing where it was scaled, outliers removed, transformed and then used for predictive modelling.

Modelling The models were be based on supervised learning. The target variable (churn) will be predicted using the feature columns. Four models were made. Hyperparameter tuning was applied and the optimal parameters were selected using the GridSearchCV. The models include: 1. Logistic Regression 2. Decision tree 3. Random Forest 4. K-Nearest Neighours

The Random forest model stood out as the top choice boasting exceptional accuracy of 92.857143% and an impressive F1 score of 86.1548%. This means it excels at precisely classifying instances while maintaining a balanced trade-off between precision and recall. It also had the highest AUC of 0.847.

Evaluation Customer Service Calls : This has the highest feature importance. This suggest that customers who frequently contant customer service may be more likely to churn. This could be a sign of unresolved issues or dissatisfaction.

Total Day Charge and Total Day Minutes: This implies that the amount a customer is charged during the day and the total minutes spent on day calls plays a crucial role in predicting churn. High charges indicate dissatisfaction with pricing.

Total Evening Minutes and Total Evening Charge: Similar to day metrics, evening minutes and charges are also relevant. Customers who spend more time on evening calls or charged more for evening calls might be prone to churning.

Number of Voicemail Messages: Presence or absence of a voicemail plan and the number of voicemail messages impact churn.

Total International Calls: Customers making a higher number of international calls could be more likely to churn, possibly due to the cost or service quality of international calls.

Voice mail plan: Whether a customer has a voicemail plan or not affects churn behavior.

Total night charge: Nighttime usage contribute to the likelihood of churn

State: The state in which a user resides can influence their propensity for churn.

Recommendations

Customer Service Calls: Improve customer service quality and response times to address customer issues promptly. Implement proactive customer service outreach to identify and resolve problems before they lead to churn.

Total Day Charge and Total Day Minutes: Offer competitive pricing and transparent billing for day calls to reduce customer dissatisfaction with charges. Create cost-effective day call packages to attract and retain price-sensitive customers.

Total Evening Minutes and Total Evening Charge: Ensure pricing for evening calls is reasonable. Consider offering evening call packages to cater to customer preferences for evening communication.

Number of Voicemail Messages: Encourage customers to use voicemail services through promotions or incentives. Ensure that voicemail services are user-friendly and add value for customers.

Total International Calls: Review international call rates and quality to make them more attractive. Offer international calling plans or discounts to retain customers who make international calls.

Voice Mail Plan: Promote the benefits of having a voicemail plan to customers. Consider bundling voicemail plans with other services to increase adoption. Total Night Charge:

Leverage the importance of nighttime usage by offering night-specific plans or benefits. Ensure that nighttime service quality meets customer expectations.

State: Analyze regional differences in churn rates and tailor marketing and retention efforts to address specific state-level issues. Monitor and improve network and service quality in regions with higher churn rates.

Conclusion

In conclusion, this project aimed to identify and understand factors contributing to customer churn in a telecommunications company. Through data analysis and predictive modeling, we discovered significant predictors of churn, including customer service calls, pricing-related metrics, voicemail usage, and international call patterns.

The Random Forest model emerged as the most suitable for predicting churn, offering high accuracy and F1 score. It demonstrated the ability to handle complex relationships between features and capture important patterns in the data. Additionally, feature importance analysis shed light on which factors significantly impact customer churn.