

kamaupaul / Churn_in_Telecom_ProjectPublic

Reducing Customer Churn using Predictive Modeling

GPL-3.0 license

0 stars0 forks

Star

Watch

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

main

...

kamaupaul

 ...

1 minute ago

View code

README.md

Churn_in_Telecom_Project

Reducing Customer Churn using Predictive Modeling

Overview



SyriaTel Mobile Telcom is a provider of mobile telecommunication and data services based in Damascus, Syria. The company offers services including calls, news, message, GSM and internet services, thereby making the life of customers easier with reasonable prices.

Telecom industry is a very competitive market, these industries like SyriaTel experience an average of 15%-25% annual churn rate. Churn has strong impact on the life time value of the customer because it affects the length of service and the future revenue of the company. Telecom companies spend hundreds of dollars to acquire a new customer and when that customer leaves, the company not only loses the future revenue from that customer but also the resources spend to acquire that customer.

To reduce churn rate the company aims to predict any potential churners in advance and take appropriate measures to ensure reduction of churn rate. This project aims to build predictive models to identify factors that contribute to churn.

When the customer has stopped using the service for a while it maybe too late to retain them, customers do not instantly decide to switch to another competitor rather they take sometime to decide, hence by detecting them before they decide we can potentially reduce the future churn rate.

Business Understanding

Customer churn refers to the phenomenon where customers or subscribers end their relationship with a company or service provider. The primary objective of any business is to reduce customer churn. By identifying potential churners in advance, SyriaTel can take precaution measures to retain these customers.

Such measures may include:

- Improved customer service.
- Addressing the issues that may lead to churn.
- Making targeted offers and,
- Targeted market campaigns among others.

Churn can have great significant financial impact on the business as high churn leads to loss of recurring revenue, damage of brand's reputation among many other effects.

This project aims to predict customer churn by developing an algorithm to predict the churn rate based on customer usage pattern.

Objectives

- To determine attributes that contribute to customer churn.
- To build a classification model that predicts customer churn.
- To achieve a recall score for the model of at least 70%
- To make valid recommendations to SyriaTel on ways they can reduce customer churn.

Data Understanding

The dataset has customer usage pattern and whether the customer has churned or not. I will develop an algorithm to predict the churn score based on usage pattern. The predictors provided are as follows:

1. "state", string. 2-letter code of the US state of customer residence
2. "account_length", Number of months the customer has been with current telco provider
3. "area_code", a string 3 digit area code.
4. "international_plan", (yes/no). The customer has international plan.
5. "voice_mail_plan", (yes/no). The customer has voice mail plan.
6. "number_vmail_messages", numerical. Number of voice-mail messages.
7. "total_day_minutes", numerical. Total minutes of day calls.

8. "total_day_calls", numerical. Total minutes of day calls.
9. "total_day_charge", numerical. Total charge of day calls.
10. "total_eve_minutes", numerical. Total minutes of evening calls.
11. "total_eve_calls", numerical. Total number of evening calls.
12. "total_eve_charge", numerical. Total charge of evening calls.
13. "total_night_minutes", numerical. Total minutes of night calls.
14. "total_night_calls", numerical. Total number of night calls.
15. "total_night_charge", numerical. Total charge of night calls.
16. "total_intl_minutes", numerical. Total minutes of international calls.
17. "total_intl_calls", numerical. Total number of international calls.
18. "total_intl_charge", numerical. Total charge of international calls
19. "number_customer_service_calls", numerical. Number of calls to customer service

Target Variable is:

- Churn: if the customer has churned (1=yes; 0 = no)

Methods

- Data exploration and cleaning.
- Data preparation and feature engineering.
- Machine Learning.

Technologies used

- Python language.
- Pandas
- Numpy
- Scikit-learn
- Matplotlib.
- Seaborn

Modeling

At this step I built three models to and evaluated their performance. I then chose the model that had the best performance and tuned it.

The models that I used are:

- 1.Logistic Regression
- 2.Decision Trees

3.Random Forest

4.Final Tuned model.

Evaluation

I used three classification models to predict churn and a more tuned model to improve the evaluation metrics i.e:

- Logistic regression which had a:
 - Train Accuracy: 65%
 - Test Accuracy: 63%
 - Train Recall: 89%
 - Test Recall: 80%
- Decision Tree which had a:
 - Train Accuracy: 95%
 - Test Accuracy: 93%
 - Train Recall: 75%
 - Test Recall: 72%
- Random Forest with a:
 - Train Accuracy: 89%
 - Test Accuracy: 88%
 - Train Recall: 23%
 - Test Recall: 12%
- Tuned Random Forest with a:
 - Train Accuracy: 94%
 - Test Accuracy: 95%
 - Train Recall: 78%
 - Test Recall: 72%

The best parameters for random forest are criterion 'entropy', Maximum depth of 6, minimum sample leaf of 3 and a minimum sample split of 10.

Findings

- A customer who had more customer service call are most likely to churn. This implies the customers who did not have a customer service call are most likely satisfied from the services offered by the company.
- A higher percentage of the customers who churned had an international plan.
- High total day minutes spent potentially leads to high churn rate, this implies that a customer was charged based on the total number of minutes spent.

- Total evening charge and night minutes also increased the churn rate.

Conclusion

The final model that I used to predict customer churn was Random Forest with the tuned hyperparameters (criterion 'entropy', Maximum depth of 6, minimum sample leaf of 3 and a minimum sample split of 10) as it has the least number of false negatives with a better recall score. The most most important features that lead to customer churn were:

- Total day minutes.
- Total day charge.
- Customer service calls.
- Internatinal plan.
- Total evening charge.

Recommendations.

1. Improve the quality of customer service provided
2. Review the international plan.
3. Review on the prices how they charge customers based on the total number of minutes.

Limitations.

- The dataset used has limited number of features while there could be other factors that contribute to customer churn.
- The classification models used in this analysis make certain assumptions, such as independence, and normality. Violations of these assumptions could impact the accuracy and reliability of the results.

Next steps.

1. New features can be generated from the existing data to provide more insights.
2. Gather additional data on customer churn that could be included in the model to improve the recall.
3. Review the models to reflect the current factors that leads to customer churn.

Releases

No releases published

[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

- Jupyter Notebook 100.0%