



 S-Kagwi / Phase-3-Project Public View license 0 stars  0 forks Star Watch Code  Issues  Pull requests  Actions  Projects  Wiki  Security  Insights  Se main ▾

...

 S-Kagwi updated README ...13 minutes ago  34[View code](#) README.md 

SyriaTel Customer Churn Analysis



Author: Samwel Kagwi

Business Understanding

SyriaTel, a telecommunications company would like to reduce money spent in the event of losing customers and acquisition of new customers. The task at hand is to use machine learning algorithms to be able to predict the likelihood of a customer churning from SyriaTel. Churning refers to when a subscriber cancels, stops or does not renew a subscription service.

Data Understanding

The utilized dataset was source from Kaggle.

Data Preparation

The following steps were taken to prepare the data:

- Checking for missing or duplicated values
- Checking the data types
- Converting columns of the object data types, to a numerical data type.
- Dropping columns that were irrelevant to this project ('area_code').

Modelling

- Baseline model: Logistic regression
- Model 2: Decision tree
- Model 3: KNN
- Model 4: Random forest

Evaluation Metric

I chose to use Recall as my evaluation metric. The recall score is true positive divided by the true positive plus the false negative. It is the measure of actual observations which are predicted correctly. I chose this metric because we want to capture as many positives as possible, and doing so would reduce the number of false negatives. In this case, a false negative would be the number of customers identified as churned, but were classified as not churned.

Model 1: Baseline Model

	precision	recall	f1-score	support
0	0.95	0.75	0.84	581
1	0.31	0.76	0.44	86
accuracy			0.75	667
macro avg	0.63	0.75	0.64	667
weighted avg	0.87	0.75	0.79	667

Results:

- Recall Score (did not churn): 75%
- Recall Score (chur): 76%

Model 2: Decision Tree

	precision	recall	f1-score	support
0	0.95	0.98	0.97	581
1	0.85	0.67	0.75	86
accuracy			0.94	667
macro avg	0.90	0.83	0.86	667
weighted avg	0.94	0.94	0.94	667

Results:

- Recall Score (did not churn): 98%
- Recall Score (churn): 67%

Model 3: KNN

	precision	recall	f1-score	support
0	0.93	0.80	0.86	581
1	0.31	0.60	0.41	86
accuracy			0.78	667
macro avg	0.62	0.70	0.64	667
weighted avg	0.85	0.78	0.80	667

Results:

- Recall Score (did not churn): 80%
- Recall Score (churn): 60%

Model 4: Random Forest

	precision	recall	f1-score	support
0	0.94	0.99	0.97	581
1	0.88	0.60	0.72	86
accuracy			0.94	667
macro avg	0.91	0.80	0.84	667
weighted avg	0.94	0.94	0.93	667

Results:

- Recall Score (did not churn): 99%
- Recall Score (churn): 60%

Analysis

The steps for modeling went as follows: build a model, print classification report and confusion matrix for model, and finally determining the fit of the model. The 1st linear regression model did not perform as well as we had expected. To fix this, we used SMOTE to balance the data, by creating synthetic values to balance our existing data. The 2nd model showed higher scores that made this model adequate enough to use if chosen. This model was the DecisionTreeClassifier. It showed high results on the classification report and promising numbers on the confusion matrix. The 3rd and 4th models plotted were KNN and Random Forest, respectively. These models did not perform as well as the 2nd model.

Recommendations

- Customers that called customer service at least 4 times have a significantly increased chance of churning. Managers should come up with new training techniques to help customer service representatives assist these disgruntled customers.
- Carry out further investigations on why customers with international plans have a higher percentage of churning.
- The firm should get more balanced data in order to improve the classification models.
- The firm should look into more variables that may have an effect on churn, such as: gender, regions and age-group.

Repository Contents

- └─ images: images used in PPT and readme
- └─ .gitignore
- └─ CONTRIBUTING.md
- └─ LICENSE.md
- └─ README.md: project information and repository structure
- └─ Working_Notebook
- └─ churn_data: data used for modeling and analysis
- └─ notebook.pdf
- └─ presentation.pdf
- └─ presentation.ppt: Non-technical Presentation for Stakeholders
- └─ student.ipynb: jupyter notebook used for modeling

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%