

SYRIATEL CUSTOMER CHURN STUDY

Gamze Turan

August 4, 2022

Content

- Overview
- Business Problem
- Data
- Models
- Evaluation Metrics
- Summary
- Future Work



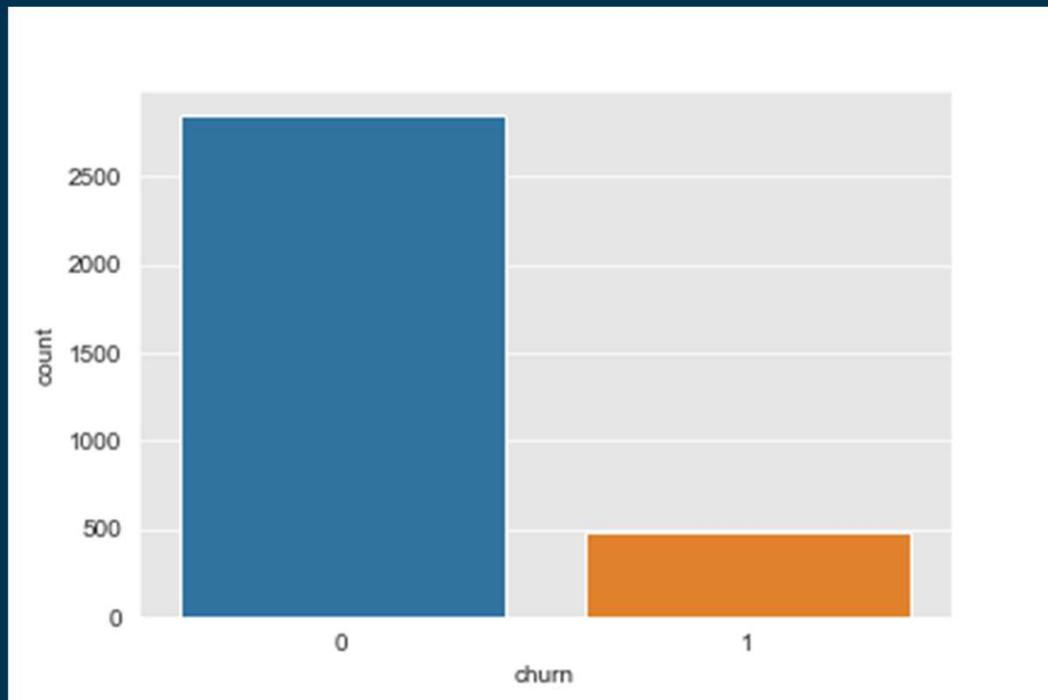
Overview

- This project is the analysis of 'SyriaTel Customer Churn' data (a telecommunication company).
- The best way to determine is to make a predictive model which will classify customers who might stop doing business with SyriaTel, using the data.
- We will build a model for classifying whether customers will stop business True or False.

BUSINESS PROBLEM

- Detecting which customers are likely to leave a service or to cancel a subscription to a service
- Select a model that will be the most accurate in predicting which clients will discontinue doing business with SyriaTel.

Data



- The data comes from SyriaTel and includes information about their customers.
- The raw data shows 95% of the customers are churn

Total Customer Spend Over Time



- You can see a line of cancelling customers above the staying ones, indicating higher spend for some cancelling customers.

Modeling

Objective : Predict churn customer

- Churn: customers or subscribers stop doing business with a company or service
- The prediction will be True(1) or False(0)
- Binary Classification used for modeling.

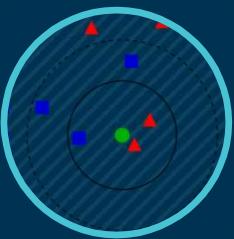
Data Processing:

- Split the data into y : target (churn), X : all predictors
- Used StandardScaler to standardized data
- Split the data into training, validation and testing subset

Several Classifiers models are tested



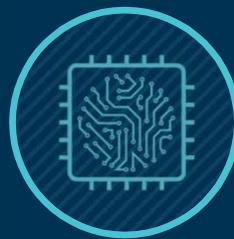
Logistic
Regression



K-Nearest
Neighbors



Decision Tree



Random Forest



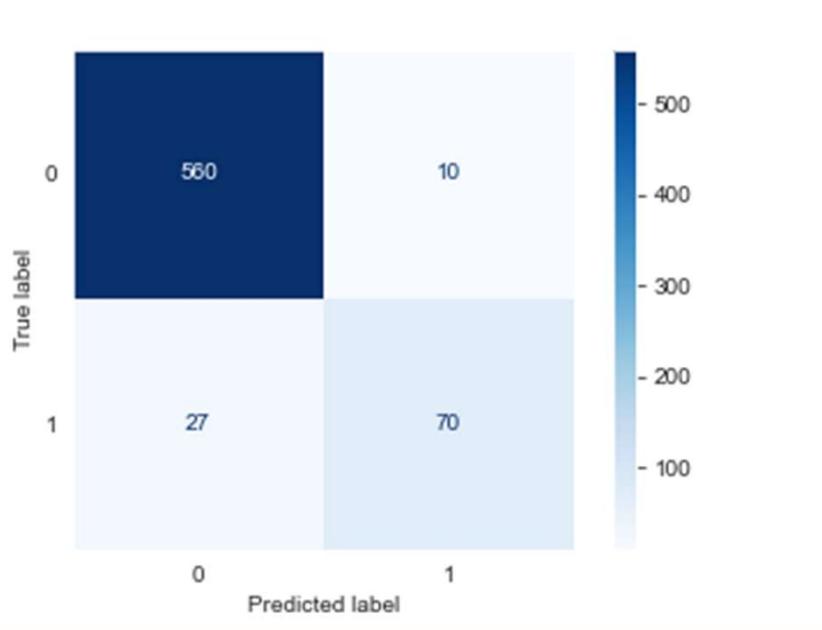
XGBoost



Evaluation Metrics Score

model	precision	recall	accuracy	f1	auc
Logistic Regression	0.328	0.691	0.750	0.445	0.790
K-Nearest Neighbor	0.276	0.515	0.733	0.360	0.658
Decision Tree	0.778	0.722	0.930	0.749	0.894
Bagging Classifier	0.645	0.825	0.909	0.724	0.918
Gradient Boost	0.772	0.732	0.930	0.751	0.923
Adaboost	0.525	0.536	0.862	0.531	0.822
Random Forest	0.741	0.619	0.913	0.674	0.917
XGBoost	0.857	0.742	0.945	0.796	0.924

Interpretation



- XGBoost Classifier performed the best
- XGBoost model improved with additional parameter tuning:

Presicion	Recall	Accuracy	F1-score
0.87	0.72	0.94	0.79

- Final model identifies the 72% churn customers correctly identify.
- 87% of the true churn customers predicted
- The Hamonic Mean of Precision and Recall is 79%

Summary

- Statistics of the Final Model in concise:
 1. It clearly identifies 74% of the real churning customer.
 2. 83% of the customers whose anticipated churn was captured by the algorithm definitely did so (clearly remember).(accuracy)
 3. The f1-score's Harmonic Mean of Precision and Recall is 78%.
- **Unique identifiers:**
 - 72 confirmed positives were found
 - There are 556 true negatives.
 - 14 false alarms were discovered.
 - 25 erroneous alarms were discovered.
 - 72 out of 125 customers who churn are successfully identified.

Recommendations:

- Improve international plan to attract customers.
- For greater satisfaction, revamp its helpdesk(customer service).
- Accept a deal at discount with enough cumulative day moments.

Next Step

- Ensure smooth functioning of the XGBT design(completed model).
- To understand how parameter influences the performance, browse for it properly
- To facilitate a better understanding and familiarity of each parameter exploited in grid search.
- Analyze the influence of additional hyperparameter.
- To evaluate performance of the model and to alter parameters, use a scaled f1 score that emphasizes recall more accuracy

Thank You

Email: ginaturan15@gmail.com

Github: [@gt1719](https://github.com/gt1719)

Linkedin:<https://www.linkedin.com/in/gamze-turan-958a8012b/>