

SyriaTel Customer Churn Study

Flatiron School

Content

- Overview
- Business Problem
- Data
- Method
- Model
- Interpret
- Future Work
- Questions



Overview

- This project is the analysis of 'SyriaTel Customer Churn' data. The SyriaTel is a telecommunication company.
- The purpose of the study is to predict whether a customer will ("soon") stop doing business with SyriaTel.
- The Data Classification models are used in analysis.

Project Details:

- GitHub: <https://github.com/kamileyagci/dsc-phase-3-project>

Business Problem

The telecommunication company, SyriaTel, hired me to analyze the Customer Churn data. The company wants to understand the customer's decision to discontinue their business with SyriaTel. The results of the analysis will be used improve the company finances.

This study will

- Search for the predictable pattern for customer decision on stop or continue doing business with SyriaTel
- Choose a model which will best identify the customers who will stop doing business with SyriaTel

Data

- The data contains 3333 entries/customers.
- For each customer, the data includes 21 types of information:
 - state, area code, phone number
 - international plan, total intl minutes, total intl calls, total intl charge
 - voice mail plan , number vmail messages
 - total day minutes, total day calls, total day charge
 - total evening minutes, total evening calls, total evening charge
 - total night minutes, total night calls, total night charge
 - customer service calls
 - account length
 - churn (= activity of customers leaving the company and discarding the services offered)

Model (1)

- Goal: Predict 'churn' value
 - churn: activity of customers leaving the company and discarding the services offered
 - The prediction will be True (1) or False (0)
 - Binary classification is used for modeling.
- Pre-processing:
 - Divided the dataset into y: target 'churn' and X: all predictors
 - Standardize the data
 - Split data into training and testing subsets:
 - Training data: used to train our model
 - Testing data: used to validate the model

Model (2)

- Several classifier models are tested:
 - Logistic Regression
 - K-Nearest Neighbors
 - Decision Trees
 - Random Forest
 - XGBoost
- Each model is optimized based on f1-score

Model (3)

- Evaluation metrics are used to measure the performance of the models:
 - precision: What percentage of model predictions are true?
 - recall: What percentage of the classes we're interested in were actually captured by the model?
 - accuracy: Out of all the predictions our model made, what percentage were correct?
 - f1-score: Harmonic Mean of Precision and Recall
- I focus on 'f1-score' and 'recall' for model comparison based on the goal

Model (5)

- Evaluation Metric Scores:

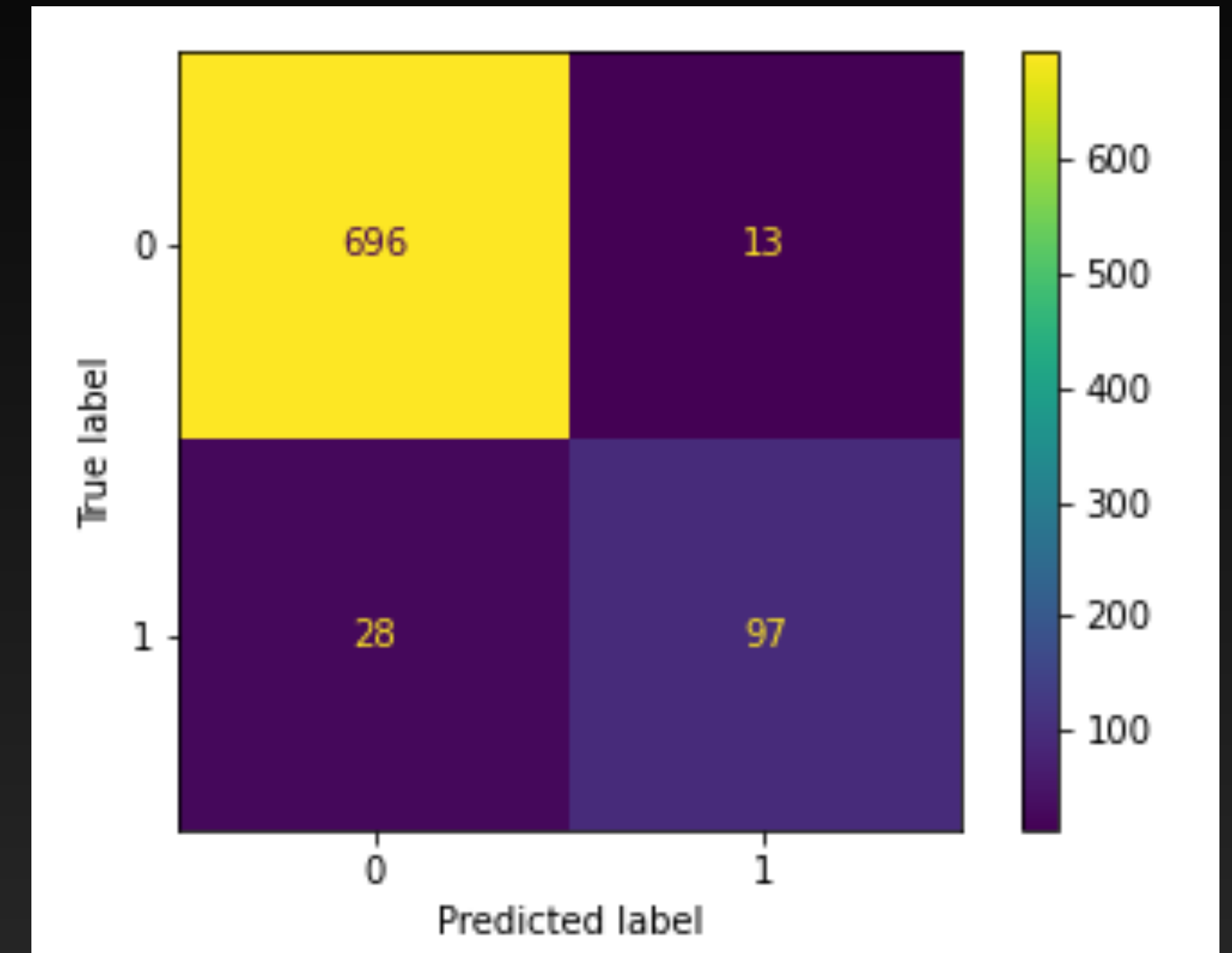
	precision	recall	accuracy	f1
model				
Logistic Regression	0.372	0.792	0.769	0.506
K-Nearest Neighbor	0.289	0.616	0.716	0.394
Decision Trees	0.682	0.704	0.906	0.693
Random Forest	0.788	0.712	0.928	0.748
XGBoost	0.887	0.752	0.948	0.814

Interpret (1)

- XGBoost Classifier Model has the best performance
- XGBoost Model is improved with additional parameter tuning:

Precision	Recall	Accuracy	f1-score
0.88	0.78	0.95	0.83

- Final model identifies the 78% of the true churn customers correctly.
- Among the model predicted churn customers, 88% of them are true churn customers.
- The Harmonic Mean of Precision and Recall (f1-score) is 83%.



On test data:

- Model identifies 97 out of 125 churn customers correctly.
- 97 out of 110 predicted churn customers are real churn.

Interpret (2)

- Top five Important Features: international plan, voice mail plan, number of voice mail messages, customer service calls, total day minutes
- **Characteristic of churn customers:**
 - The churn customers are **more likely to have international plan** than continuous customers.
 - The churn customers are **less likely to have voice mail** plan than continuous customers.
 - The churn customers have **less voice mail messages** than continuous customers (as a result of less voice mail plan)
 - The churn customers have **more customer service calls** than continuous customers.
 - The churn customers have **more total day minutes** than continuous customers

Business Recommendations

- Improve the international plan to keep/attract customers
- Improve customer service center for better customer satisfaction
- Offer a discounted plan with high total day minutes

Future Work

- Improve the XGBoost model (final model) performance
 - Search each parameter separately to understand the effect on performance
 - Obtain a more sensitive/informed parameter range to be used in grid search
 - Study the effect of other hyperparameters
- Use weighted f1-score, with more weight on recall than precision
 - to compare model performance
 - and for parameter tuning

Questions?