

Module 3 final project

Zhiqiang Sun

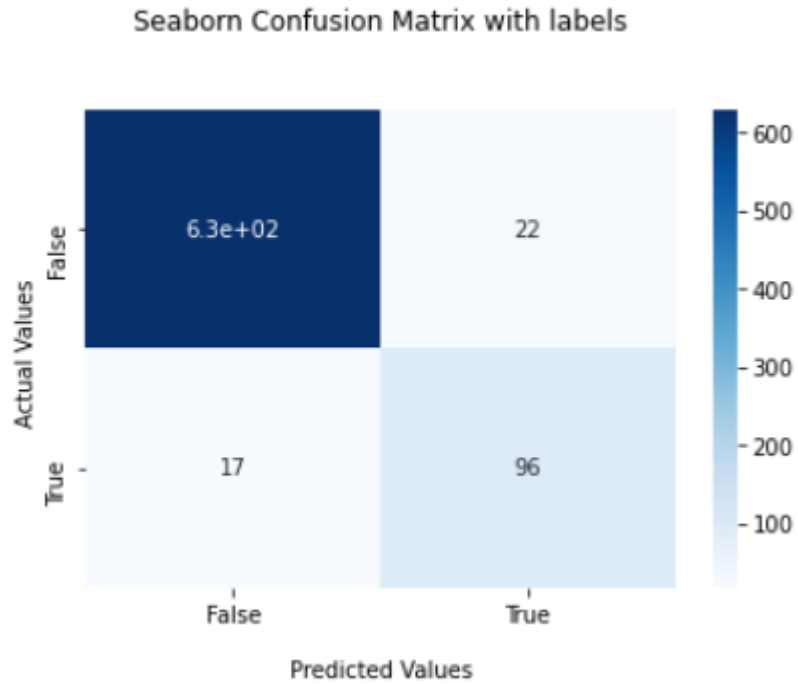
- Business Problem
- SyriaTel is a telecommunications company. They have a series of data contains some basic information of their customers and whether they will continue or stop doing business with them.
- They want us to make a model with the dataset and predict whether a customer will (“soon”) stop doing business with them. Based on the model, we can find some potential suggestion to keep customers with their business.

Plan

- Since the SyriaTel Customer Churn is a binary classification problem , I will try to use several different algorithms to fit the data and select one of the best one. The algorithms I will try include Logistic Regression, k-Nearest Neighbors, Decision Trees, Random Forest, Support Vector Machine.
- The criteria to selected model will be f1 score and confusion matrix.

Results

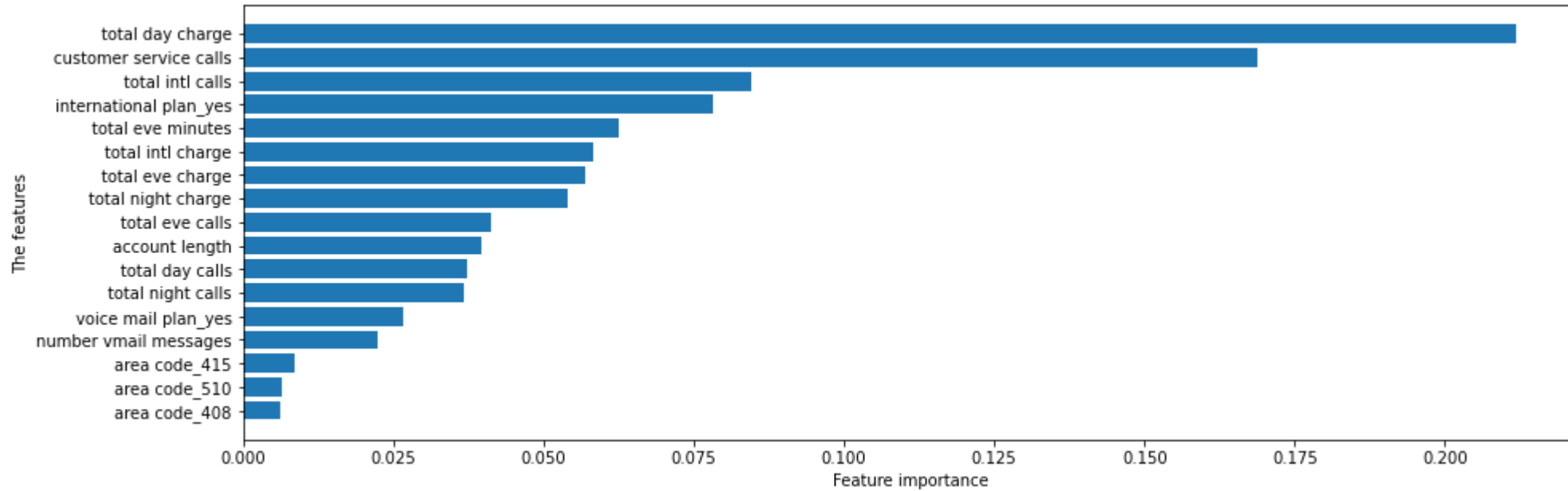
The final selected model is Random Forest Classifier.



accuracy_score is 0.94895
f1_score is 0.83117
recall_score 0.84956
precision_score 0.81356

Compared to other four models, the model of Random Forest gives us best results on the f1 scores. The TP in confusion matrix increased and the FP, FN decreased. Thus, we selected random Forest for our final model.

The feature importance:



The top 5 important features to affect customers decisions.

total day charge 0.21184

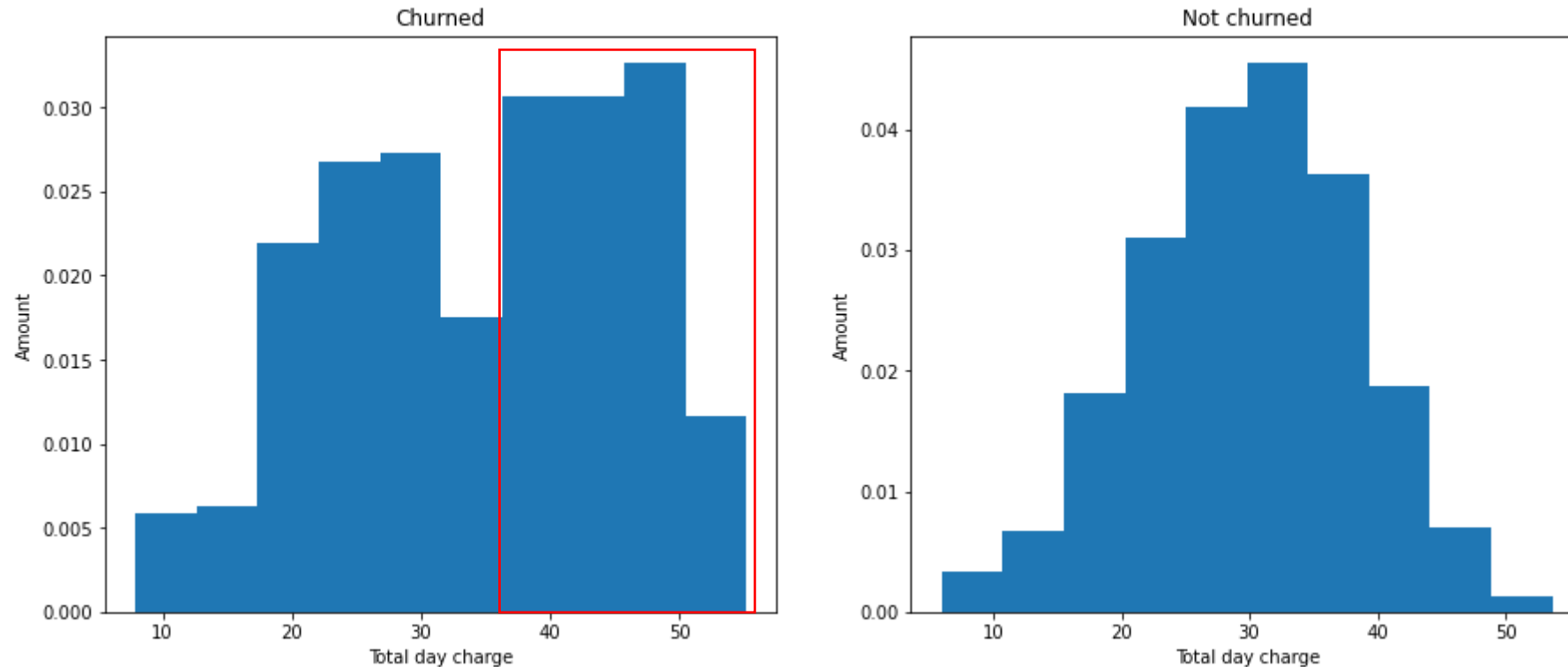
customer service calls 0.16868

total intl calls 0.084634

international plan_yes 0.0781320

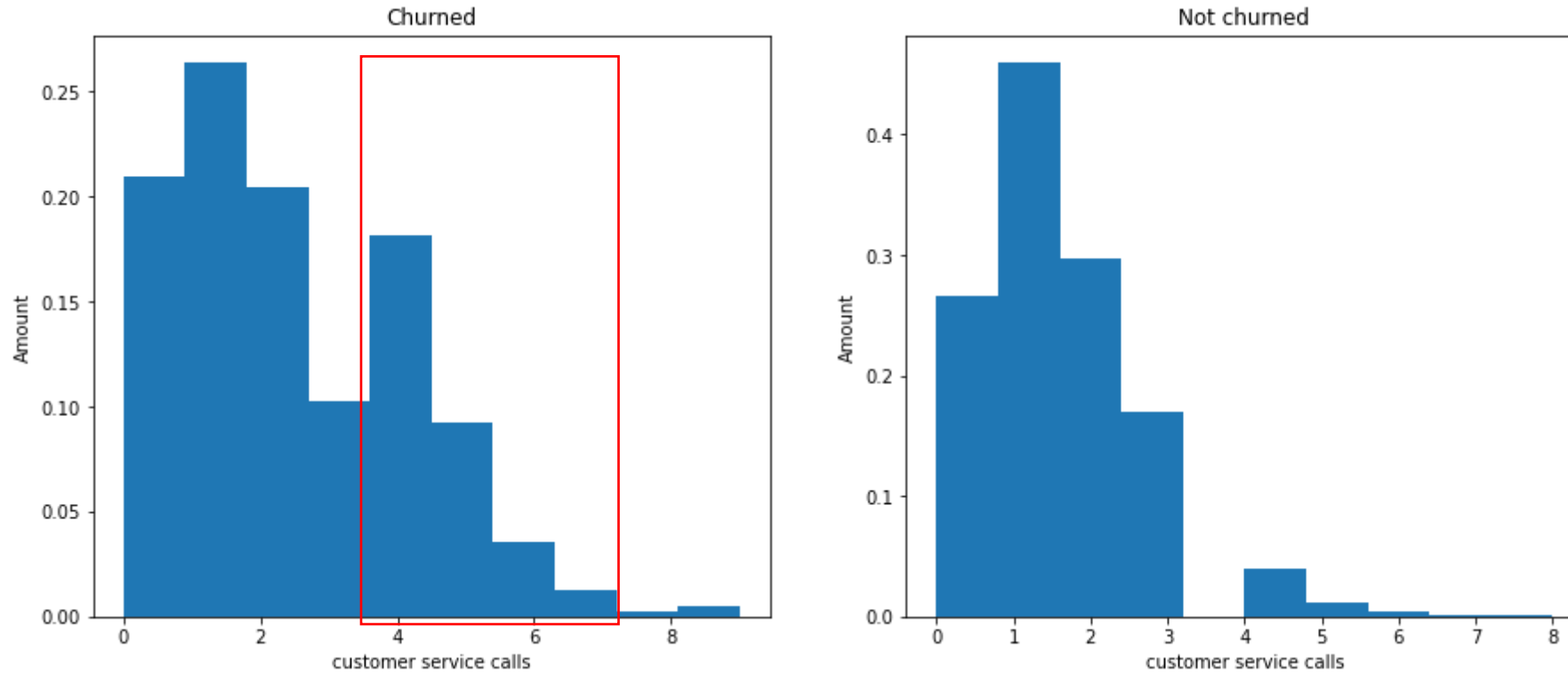
total eve minutes 0.06242784

The affections of Total day charge.



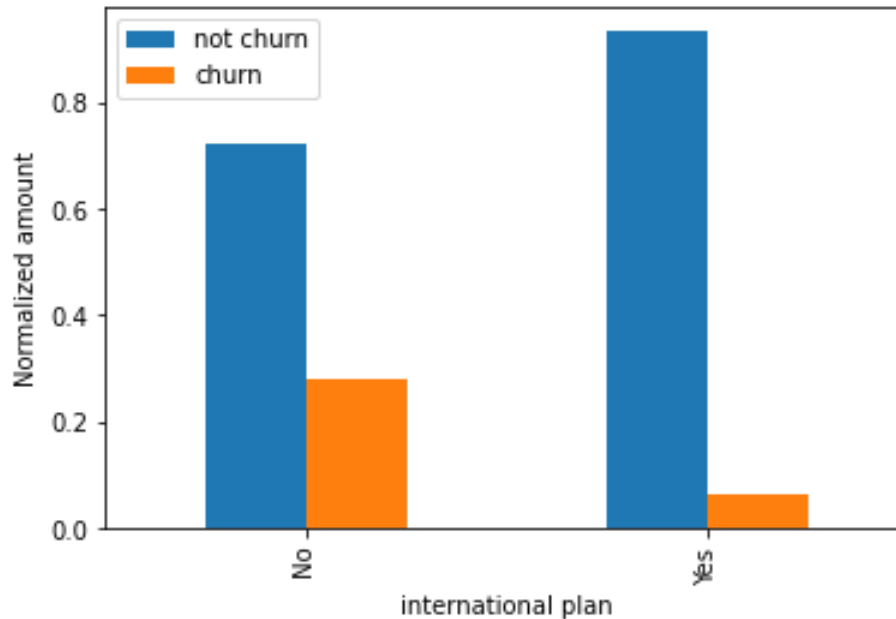
The histograms for customers who churned and not churned show that the total day charge have a lot of overlap with each other. The customers who had total day charge more than 40 have more chance to churn the plan.

The affections of the number of customer service calls.



The histogram are similar to each other. However, the customer who had four or more times international calls had higher chance to churn the plan.

The affections of customers who have international plan or not.



	not churn	churn
No	0.721198	0.278802
Yes	0.933944	0.066056

This data show that the customers who have international plan have 27% chance to churn the service. But the customers who do not have the international plan have only 6.7% chance to churn the service.

Conclusion

- We polished our original data by removing the outlier and catalyzing the necessary columns. We then tested several models to fit out data and selected the best one, Random Forest. The final score of predicting is 0.83117, which is very high.
- By plotting the feature importance, we found that the top 5 weighted features are total day charge, customer service calls, international plan_yes, total intl calls, and eve minutes. We then plot the histogram of each feature separated by the customers who churn and not churn the plan.
- We found that customers who had day charge more than 40 or had customer service call four and more or had an international plan had a higher chance of churning the service.
- The results of total day charge and customer service calls suggest customers using Syriatel service more often have a higher chance of churning the service. Thus, Syriatel company can promote the service charge to attract people to use more of that. They also need to make the customer service friendlier and more professional to help customers address the problems.
- The international plan is also crucial for customers to churn the service. Based on this point, Syriatel can also make some memorable plans if more customers use the international program.

THANKS!

- Zhiqiang Sun
- sunzhiqiang04@gmail.com