CUSTOMER CHURN PREDICTION

by

DHANUSH BEDRE

30-06-2022

Abstract

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customers to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main objective is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques and builds a new way of feature engineering and selection. The model was prepared and tested by working on a large dataset that was taken from Kaggle.com. The dataset contained about 7000+ customers' information and was used to train, test, and evaluate. The model experimented with three algorithms: Logistic Regression, Random Forest, and SVC. However, the best results (after analyzing performance attributes) were obtained by applying the SVC algorithm. This algorithm was used for classification in this churn predictive model.

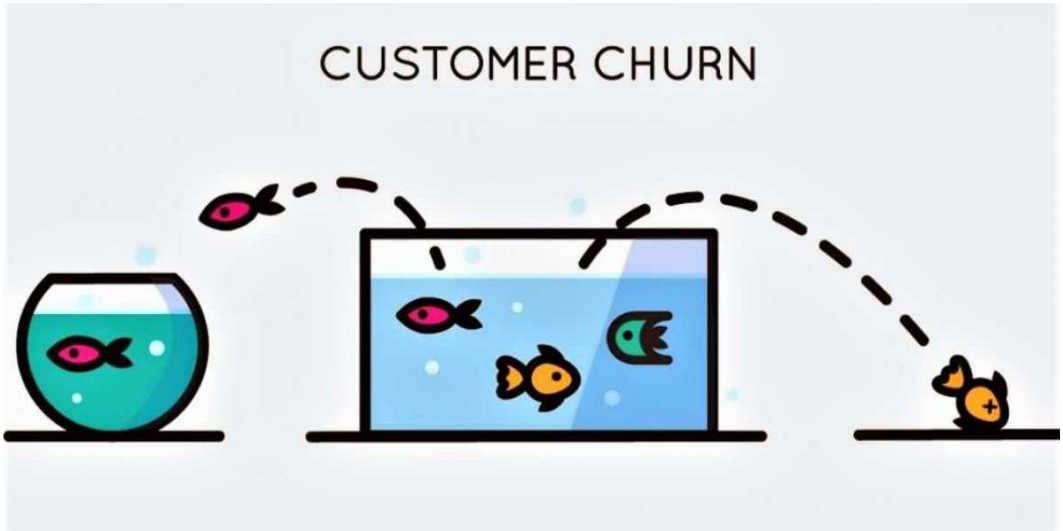
1.0 Problem Definition

The key challenge is to predict if an individual customer will churn or not. To accomplish that, machine learning models are trained based on (80-85%) of the sample data. The remaining (15-20%) are used to apply the trained models and assess their predictive power with regards to “churn / not churn”. And to demystify which features actually shows greater impact on customer churn. That information can be used to identify customer pain points and resolve them by providing goodies to make customers stay.

To compare models and select the best for this task, the accuracy & other performance scores are taken in to the consideration. Based on other characteristics of the data, for example the balance between classes (number of “churners” vs. “non-churners” in data set) further metrics are also included if needed.

2.0 Market Need Assessment

The telecommunications sector has become one of the main industries in each country. The technical progress and the increasing number of operators, as well services, raised the level of competition. Companies in this sector are working hard to compete and to upgrade their revenue. Every telco company generally follows the mentioned 3 strategies (1) Acquire new customers, (2) Upsell the existing customers, and (3) Increase the retention period of customers. These days it is a challenging task to attract new customers and at the same time to avoid contract terminations to grow their revenue generating base. Looking at churn, different reasons trigger customers to terminate their contracts, for example better price offers, more interesting packages, bad service experiences or change of customer’s personal situations. Churn analytics provides valuable capabilities to predict customer churn and also define the underlying reasons that drive it. The churn metric is mostly shown as the percentage of customers that cancel a product or service within a given period (mostly months).



3.0 Target Specification and Characterization

There are direct and indirect targets for this developed model. The direct targets include all the existing telecom companies and upcoming new telecom companies. It would help them to know the most required customer needs that lead to churn so that they can invest more potential & revenue in those areas and increase/retain their customers. And the indirect targets are Private internet service and DTH providers. These days they are providing Internet, in addition to that DTH and other services (such as OTT platforms).

4.0 External Search

The sources that I have used as a reference for analyzing the customer needs and strategies of the telecom industry to increase/retain their customers have mentioned below :

https://www.kaggle.com/code/casper6290/telcocustomerchurn/data

https://www.forbes.com/sites/forbesbusinesscouncil/2021/10/12/how-telecom-companies-can-address-customer-needs-and-stay-competitive/?sh=5e2b36146e40

https://www.intraway.com/blog/10-sales-strategies-for-the-telecommunications-industry/

https://www.axigen.com/articles/improve-customer-retention-telecom-industry_94.html#:~:text=Gather%20Customer%20Feedback%20Regularly,points%20along%20the%20customer%20journey.

5.0 Bench marking

Large telecom companies like Reliance Jio, Airtel, and Vi have been using ML models to perform Customer Churn Analysis and Classification, which identifies the needs of customers that lead to churn and uses this information to increase potential & revenue in those areas which lead to churn. But this technique would also be beneficial when applied to the small telecom companies like Quadrant, MTNL, Telenor, and BSNL since most of the needs and services are still being provided by these.

6.0 Applicable Patents

Predicting customer churn in a telecommunications network environment (US20150310336A1)

7.0 Applicable Constraints

- Data Collection (Feedback) from customers and metadata.
- Continuous data collection and maintenance

8.0 Business Opportunity

Since a similar path has been used by large companies, this can be extended to small companies, not only telecom companies but also to the local Internet & DTH service providers. Therefore, there is a fair chance of this service being a great business opportunity. Every small business that depends on sales can and would want to opt for using this service in order to always know what their customers want. The emergence of every small business is thus a fairly great business opportunity for the service provided by us.

9.0 Concept Generation & Development

In this classification problem, a basic machine learning pipeline based on a real time data set from Kaggle is build and performance of different model types is compared. The pipeline used for this example consists of 7 steps:

Step 1: Data Collection
Step 2: Exploratory Data Analysis (EDA)
Step 3: Feature Engineering
Step 4: Train/Test Split
Step 5: Model Evaluation Metrics Definition
Step 6: Model Selection, Training, Prediction and Assessment
Step 7: Hyperparameter Tuning/Model Improvement

```
df = pd.read_csv("./Telco_Customer_Churn.csv")
print(df.shape)
df.sample(5)
```

```
(7043, 21)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechS
1023	7460-ITWWP	Female	1	Yes	No	45	Yes	No	Fiber optic	No	...	Yes	
210	7841-TZDMQ	Male	0	No	No	2	Yes	No	DSL	No	...	No	
5713	8050-DVOJX	Male	1	No	No	49	Yes	Yes	DSL	Yes	...	Yes	
5749	3349-ANQNH	Female	1	No	No	59	Yes	Yes	Fiber optic	No	...	Yes	
103	5386-THSLQ	Female	1	Yes	No	66	No	No phone service	DSL	No	...	Yes	

5 rows × 21 columns

```
df.columns
```

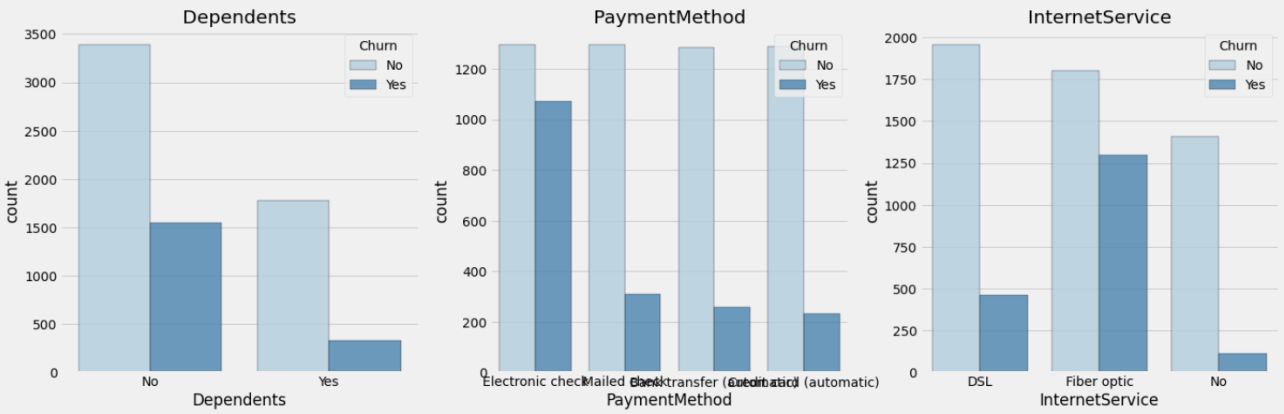
```
Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
      dtype='object')
```

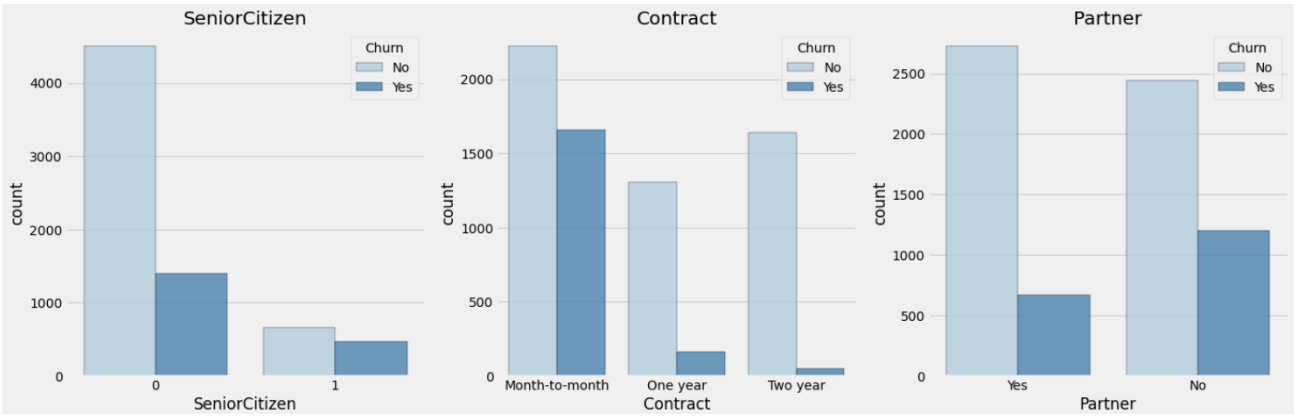
Classification labels:

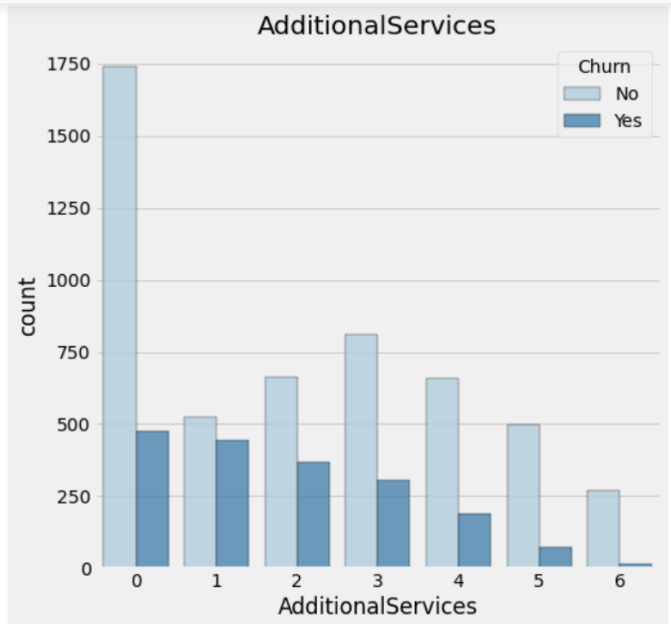
Churn — Whether the customer churned or not (Yes or No)

Customer services:
Phone Service — Whether the customer has a phone service (Yes, No)
Multiple Lines — Whether the customer has multiple lines (Yes, No, No phone service)
Internet Service — Customer's internet service provider (DSL, Fiber optic, No)
Online Security — Whether the customer has online security (Yes, No, No internet service)
Online Backup — Whether the customer has online backup (Yes, No, No internet service)
Device Protection — Whether the customer has device protection (Yes, No, No internet service)
Tech Support — Whether the customer has tech support (Yes, No, No internet service)
Streaming TV — Whether the customer has streaming TV (Yes, No, No internet service)
Streaming Movies — Whether the customer has streaming movies (Yes, No, No internet service)

Customer account information:
Tenure — Number of months the customer has stayed with the company
Contract — The contract term of the customer (Month-to-month, One year, Two year)
Paperless Billing — Whether the customer has paperless billing (Yes, No)
Payment Method — The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
Monthly Charges — The amount charged to the customer monthly
Total Charges — The total amount charged to the customer
Customers demographic info
Customer ID — Customer ID
Gender — Whether the customer is a male or a female
Senior Citizen — Whether the customer is a senior citizen or not (1, 0)
Partner — Whether the customer has a partner or not (Yes, No)
Dependents — Whether the customer has dependents or not (Yes, No)







0 – No Additional services

1 – Online Security

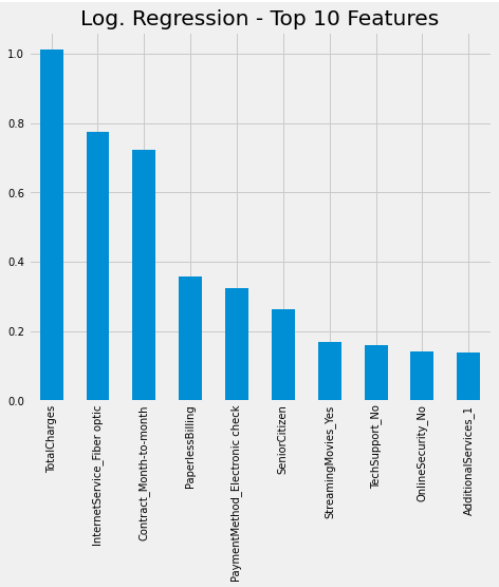
2 – Device Protection

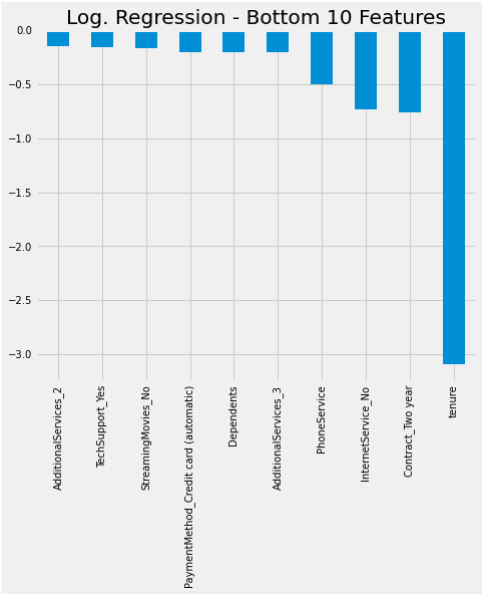
3 – Streaming Movies

4 – Tech Support

5 – Streaming TV

6 – Online Backup






```
model_params = {
    'svm': {
        'model': SVC(gamma='auto'),
        'params' : {
            'C': [1,10,20],
            'kernel': ['rbf','linear']
        }
    },
    'random_forest': {
        'model': RandomForestClassifier(),
        'params' : {
            'n_estimators': [1,5,10],
            'criterion' : ["gini", "entropy"]
        }
    },
    'logistic_regression' : {
        'model': LogisticRegression(solver='liblinear',multi_class='auto'),
        'params': {
            'C': [1,5,10]
        }
    }
}
```

	model	best_score	best_params
0	svm	0.799911	{'C': 1, 'kernel': 'linear'}
1	random_forest	0.779434	{'criterion': 'gini', 'n_estimators': 10}
2	logistic_regression	0.805032	{'C': 10}

10.0 Product Details

The final product is a day-to-day or weekly analysis (a limited period) that provides the reasons for the customer churn in a recent period of time with detailed information on what made them churn and other similar useful insights into how to increase/retain their customers and upgrade their revenue. This product requires continuous data feeding for more predictive power.

11.0 Code Implementation

https://github.com/Dhanush-22/Feynn_Labs

12.0 Conclusion

The importance of this thesis is to help telecom companies make more profit and supportive guidance from features for the small telecom companies. It has become known that predicting churn is one of the most important sources of income for telecom companies. Hence, this thesis aimed to build a system that predicts the churn of customers. To test and train the model, the sample data is divided into 80% for training and 20% for testing. I have applied feature engineering, effective feature transformation, and selection approach to make the features ready for machine learning algorithms and performed cross-validation with 10-folds for validation and hyperparameter optimization.