

TELECOM CUSTOMER CHURN PREDICTION

Using Machine Learning(Random Forest Classifier)

Developed by:Srija Dodda,Ramya Adi,Swetha Bekkanti,Prashanthi Peddireddy

Smart Bridge-Remote Summer Internship Program

1.INTRODUCTION

Churn prediction is done using predictive modeling. This is a type of ML algorithm that is generally developed in three steps. Initially, historical customer data that include information about churned customers and retained customers are collected. Once the data are prepared, a predictive model is built using R, Python or similar languages. Finally, churn predictions are made along with its confidence interval. The customer churn, also known as customer attrition, refers to the phenomenon whereby a customer leaves a service provider.

The telecommunications sector has become one of the main industries in developed countries. The technical progress and the increasing number of operators raised the level of competition. Companies are working hard to survive in this competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues: acquire new customers, upsell the existing customers, and increase the retention period of customers. However, comparing these strategies taking the value of return on investment of each into account has shown that the third strategy is the most profitable strategy. It proves that retaining an existing customer costs much lower than acquiring a new one in addition to being considered much easier than the upselling strategy. To apply the third strategy, companies have to decrease the potential of customer's churn, known as "the customer movement from one provider to another". 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 customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. Customer Churn is a key challenge in competitive markets and is highly observed in telecommunication section. Customer churns are those specific customers who have decided to leave the use of service, product, or even company and shifting to next competitor in the market. Literature reveals the following types of customer churns

- Volunteer: When customers want to quit the contract and move to the next service provider.
- Non-Volunteer: When the company quit the service to a customer.

- Silent Churn: Those customers who discontinued the contract without prior knowledge of both parties (customer-company).

1.1 overview:

Nowadays, Customer Churn has become highly important for companies because of increasing competition among companies, increased importance of marketing strategies and conscious behavior of customers in the recent years. Customers can easily trend toward alternative services. Companies must develop various strategies to prevent these possible trends, depending on the services they provide.

During the estimation of possible churns, data from the previous churns might be used. An efficient churn predictive model benefits companies in many ways. Early identification of customers likely to leave may help to build cost effective ways in marketing strategies. Customer retention campaigns might be limited to selected customers but it should cover most of the customer. Incorrect predictions could result in a company losing profits because of the discounts offered to continuous subscribers. Therefore, the right predictions of the churn customers has become highly important for the companies.

1.2 Purpose

Our aim from the project is to make use of pandas, matplotlib, render_template, request, pickle libraries from python to extract the libraries for machine learning for the customer churn prediction. Secondly, to learn how to hyper tune the parameters using grid search cross validation for the random forest classifier machine learning algorithm. And in the end, to predict whether the customer can churn the company or not ensemble techniques of combining the predictions from multiple machine learning classifier algorithms and predicting the output.

2. LITERATURE SURVEY

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data use it learn for themselves. Machine learning uses two types of techniques.

- **supervised learning:** which trains a model on known input and output data so that it can predict future outputs.
- **unsupervised learning:** which finds hidden patterns or intrinsic structures in input data.

The Machine Learning process starts with inputting training data into the selected algorithm. Training data being known or unknown data to develop the final Machine Learning algorithm. The type of training data input does impact the algorithm, and that concept will be covered further momentarily. To test whether this algorithm works correctly, new input data is fed into the Machine Learning algorithm. The prediction and results are then checked.

2.1 Existing Problem

There are a multitude of issues that can lead customers to leave a business, but there are a few that are considered to be the leading causes of customer churn. The poor customer service that nearly nine out of ten customers have abandoned a business due to a poor experience. And the other problems are a lack of ongoing customer success, natural causes that occur for all businesses from time to time, a lack of value, low-quality communications, and a lack of brand loyalty.

2.2 Proposed Solution

Machine Learning(Random Forest Classifier)

The prediction of Customer churn is done by using Random Forest. The following framework is composed of the following phases:

- ⇒ DataCollection.
- ⇒ DataPreprocessing.
- ⇒ Model Bulding.
- ⇒ Training and testing the model.
- ⇒ Getting the highest acuuracy.
- ⇒ Using trained model for prediction.
- ⇒ Application Bulding.

Random Forest is efficient for both classification and regression and the ensemble tree -based learning algorithm .Classifier is a set of decision trees from randomly selected subset of training set.It is one of the most accurate learning algorithms available .For many datasets,it provides a high accurate classifier.Classification is Principally done by making predictions based on known sample data that has been learned from traning data.

And also we have created an UI using Flask for the churn status prediction,this UI will allow the users to predict the churn status very easily and the user interface is user friendly not at least one complication in using the interface,and it can be used just by

entering some necessary details into the UI in real time it'll give the predicted value like customer is remain in the company or exit the company.

3.THEORETICAL ANALYSIS

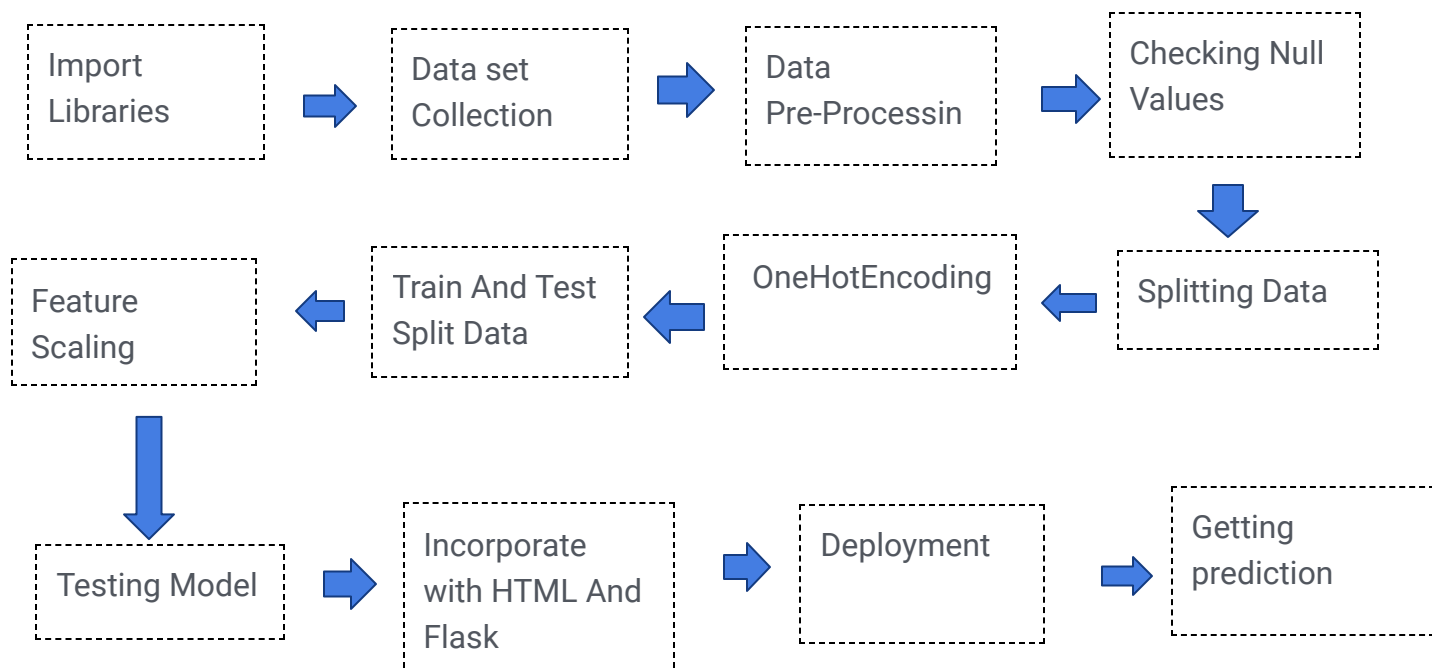
While selecting the algorithm that gives an accuracy prediction we gone through lot of algorithms which give the results high accurate and from them we selected only one algorithm for prediction problem that is Random Forest Classifier, it assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

The peculiarity of this problem is collecting the customers details real time and working with the prediction at the same time,so we can know the customers you will be remain and will exit.Accuracy is defined as the defined as the ratio of samples correctly classified by the classifier to the total number of samples for given test data set.The formula is as follows

$$\text{Accuracy} = \frac{TP+TN}{TN+FT+FN}$$

At first we got like lot of worst accuracies because we tried lot of algorithms for the best accurate algorithm,finally after all of that we tried the best suitable algorithm which gives the prediction accurately is Random Forest Classifier.And developed prediction problem for the customer churn.

3.1 Block Diagram



3.2 Software Designing

- Jupyter Notebook Environment
- Spyder Ide
- Machine Learning Algorithms
- Python(Pandas,numpy,matplotlib,seaborn,sklearn)
- HTML
- Flask

We developed this churn prediction by using the Python language which is a interpreted and high level programming language and using the Machine Learning algorithms.For coding we used the Jupyter Notebook environment of the Anaconda distributiond and the Spyder,it is an integrated scientific programming in the python language.

For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python.It is classified as a microframework because it does not require particular tools or libraries.It has no database abstraction layer,from validation,or any other components where pre-existing third-party libraries provide common functions,and a scripting language to create a webpage is HTML by creating the templates to use in the functions of the Flask and HTML.

4.EXPERIMENTAL INVESTIGATION

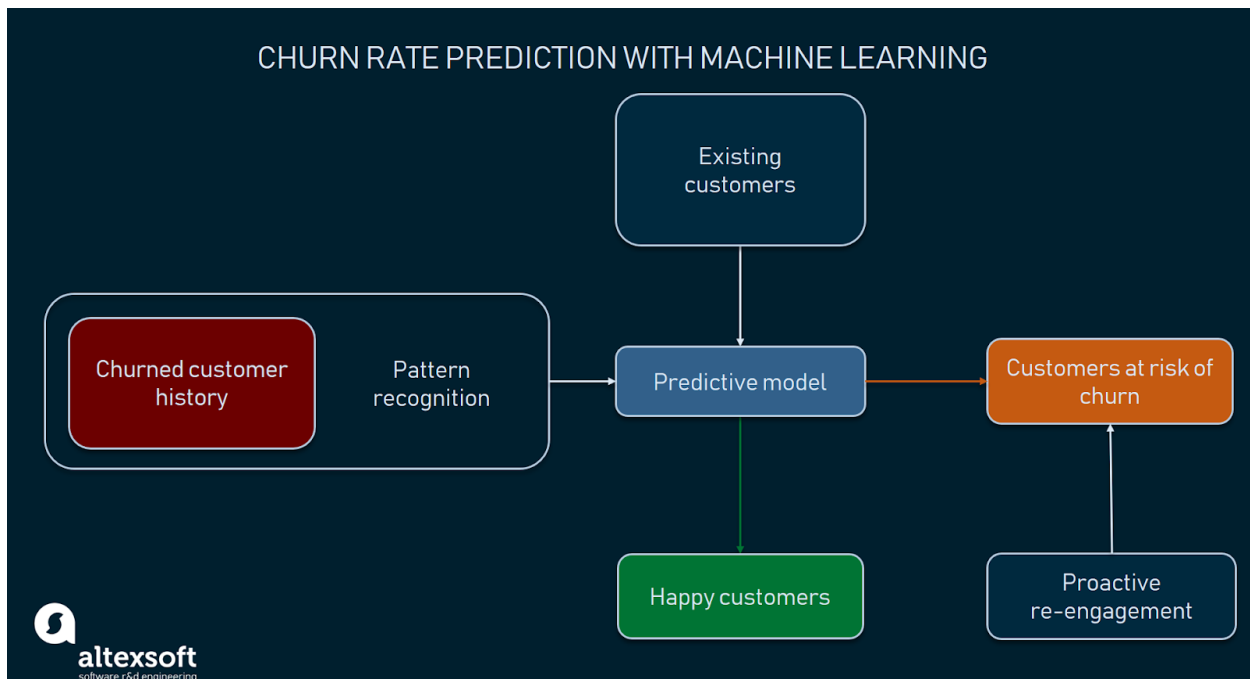
In this page the datasetwe used is derived from <https://www.kaggle.com/shrutimechlearn/churn-modelling> it contains more than 1000 original customer data of users with 21 attributes.Ater that,the missing values are filled in by means of modeinterpolation,and theduplicate or meaningless attributes are deleted, finally we have retained to 13 attributes.Those attributes were shown below in the screenshot of the datadet we used.

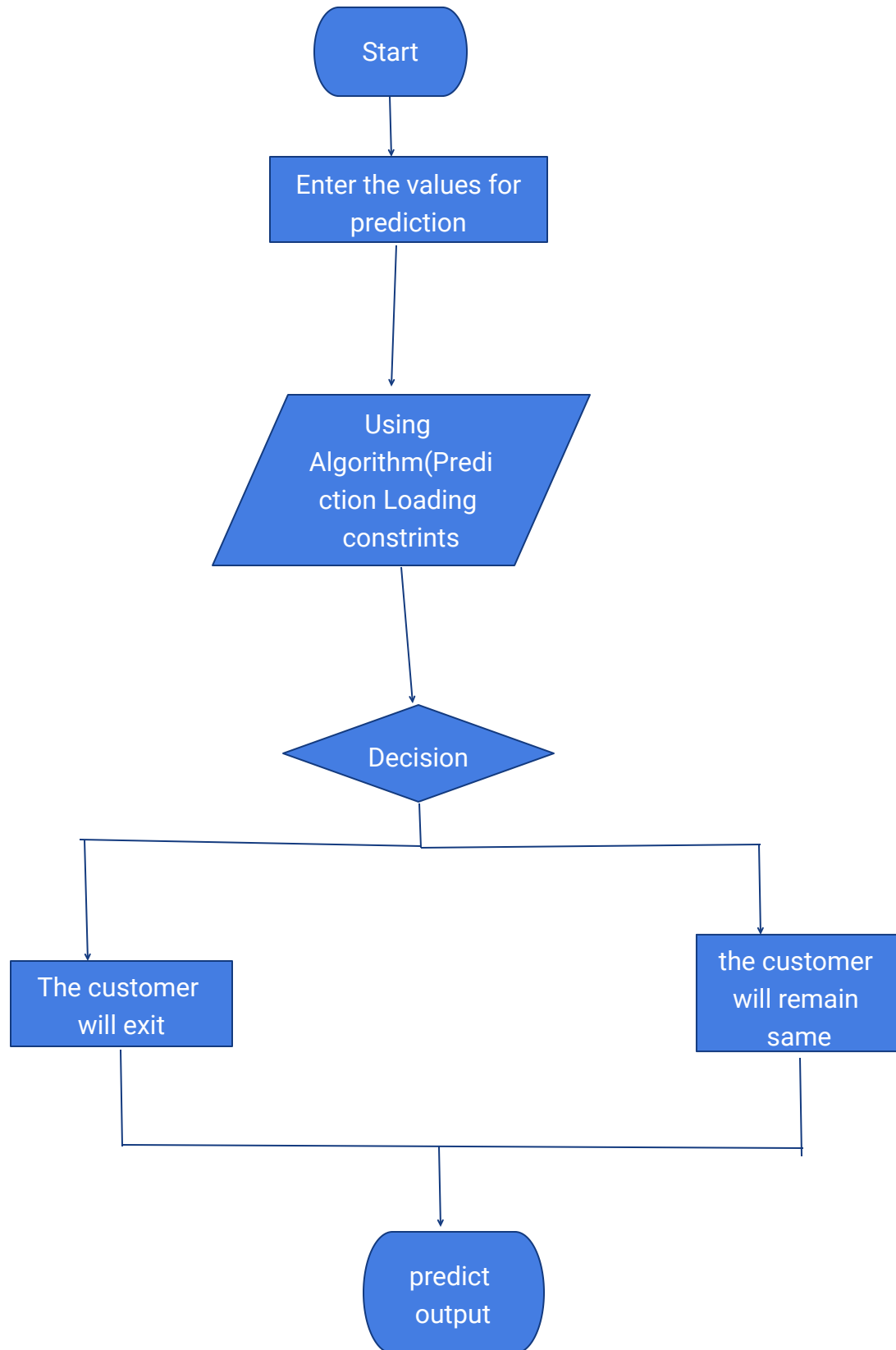
POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	SeniorCltz	Partner	tenure	PhoneServ	MultipleLir	InternetSe	TechSupp	Streaming	Streaming	PaymentM	MonthlyCl	TotalCharg	Churn										
2	0	Yes	1	No	No phone	DSL	No	No	No	Electronic	29.85	29.85	No										
3	0	No	34	Yes	No	DSL	No	No	No	Mailed che	56.95	1889.5	No										
4	0	No	2	Yes	No	DSL	No	No	No	Mailed che	53.85	108.15	Yes										
5	0	No	45	No	No phone	DSL	Yes	No	No	Bank trans	42.3	1840.75	No										
6	0	No	2	Yes	No	Fiber optic	No	No	No	Electronic	70.7	151.65	Yes										
7	0	No	8	Yes	Yes	Fiber optic	No	Yes	Yes	Electronic	99.65	820.5	Yes										
8	0	No	22	Yes	Yes	Fiber optic	No	Yes	No	Credit car	89.1	1949.4	No										
9	0	No	10	No	No phone	DSL	No	No	No	Mailed che	29.75	301.9	No										
10	0	Yes	28	Yes	Yes	Fiber optic	Yes	Yes	Yes	Electronic	104.8	3046.05	Yes										
11	0	No	62	Yes	No	DSL	No	No	No	Bank trans	56.15	3487.95	No										
12	0	Yes	13	Yes	No	DSL	No	No	No	Mailed che	49.95	587.45	No										
13	0	No	16	Yes	No	No	No interne	No interne	No interne	Credit car	18.95	326.8	No										
14	0	Yes	58	Yes	Yes	Fiber optic	No	Yes	Yes	Credit car	100.35	5681.1	No										
15	0	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	Bank trans	103.7	5036.3	Yes										
16	0	No	25	Yes	No	Fiber optic	Yes	Yes	Yes	Electronic	105.5	2686.05	No										
17	0	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Credit car	113.25	7895.15	No										
18	0	No	52	Yes	No	No	No interne	No interne	No interne	Mailed che	20.65	1022.95	No										
19	0	No	71	Yes	Yes	Fiber optic	No	Yes	Yes	Bank trans	106.7	7382.25	No										
20	0	Yes	10	Yes	No	DSL	Yes	No	No	Credit car	55.2	528.35	Yes										
21	0	No	21	Yes	No	Fiber optic	No	No	Yes	Electronic	90.05	1862.9	No										
22	1	No	1	No	No phone	DSL	No	No	Yes	Electronic	39.65	39.65	Yes										
23	0	Yes	12	Yes	No	No	No interne	No interne	No interne	Bank trans	19.8	202.25	No										
24	0	No	1	Yes	No	No	No interne	No interne	No interne	Mailed che	20.15	20.15	Yes										
25	0	Yes	58	Yes	Yes	DSL	Yes	No	No	Credit car	59.9	3505.1	No										
26	0	Yes	49	Yes	No	DSL	Yes	No	No	Credit car	59.6	2970.3	No										
27	0	No	30	Yes	No	DSL	No	No	No	Bank trans	55.3	1530.6	No										
28	0	Yes	47	Yes	Yes	Fiber optic	No	Yes	Yes	Electronic	99.35	4749.15	Yes										
29	0	Yes	1	No	No phone	DSL	No	No	No	Electronic	30.2	30.2	Yes										
30	0	Yes	72	Yes	Yes	DSL	Yes	Yes	Yes	Credit car	90.25	6369.45	No										
31	0	No	17	Yes	No	DSL	No	Yes	Yes	Mailed che	64.7	1093.1	Yes										
32	1	Yes	71	Yes	Yes	Fiber optic	Yes	No	No	Credit car	96.35	6766.95	No										
33	1	Yes	2	Yes	No	Fiber optic	No	Yes	Yes	Credit car	95.5	181.65	No										
34	0	Yes	27	Yes	No	DSL	Yes	No	No	Mailed che	66.15	1874.45	No										
35	0	No	1	Yes	No	No	No interne	No interne	No interne	Bank trans	20.2	20.2	No										
36	1	No	1	Yes	No	DSL	No	No	No	Bank trans	45.25	45.25	No										

Telco-Customer-Churn

5.FLOWCHART

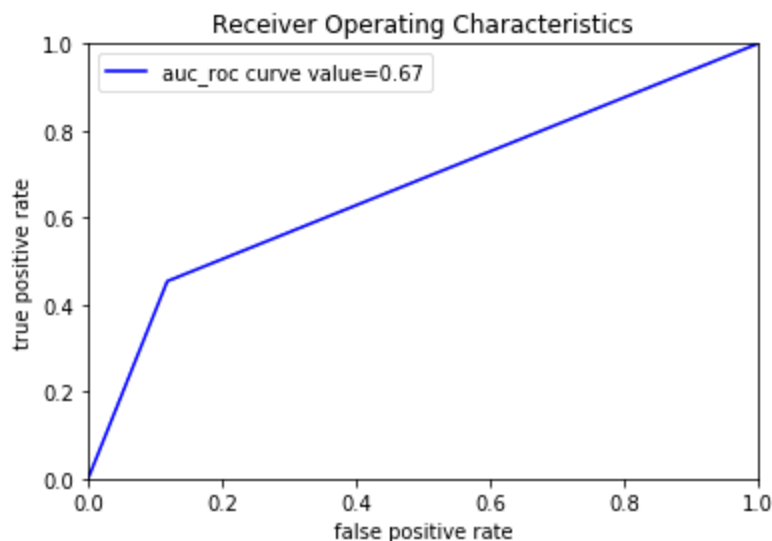




6.RESULT

In this paper, the Naive Bayes algorithm is used to predict its performance, and compared with another six machine learning methods namely the decision tree, the logistic regression, KNN, Random Forest and the SVM. The obtained results are displayed in below. The results show that, the performance of KNN and Random Forest have comparable performance than that of logistic regression, Navi bayes, SVM and decision tree, but the Random Forest still performs the best, with an accuracy of 77%, higher than the KNN with an accuracy of 70%. The ROC curve of the prediction model based on Naive Bayes are all above 0.50 , indicating that the model has strong ability of generalization.

The ROC curve of the Random Forest is shown below, the curve have only 0.50 as auc but the accuracy of this model is best out of the six algorithms at 77%.The point on the ROC curve closest to the upper left corner is the best threshold with the least classification errors, and the total number of false positive examples and false negative examples is the lowest, but our ROC stands at middle as a perfect slope. The given pie shows whose the people likely to remain in company with the data set we've took for the prediction.



The prediction Output is given below

← → ↻ ⓘ localhost:5000/login

Telecom Customer Churn Prediction

SeniorCitizen
Enter SeniorCitizen

Partner
Enter Partner

tenure
Enter tenure

PhoneService
Enter PhoneService

MultipleLines
Enter MultipleLines

InternetService
Enter InternetService

← → ↻ ⓘ localhost:5000/login

Enter InternetService

TechSupport
Enter TechSupport

StreamingTV
Enter StreamingTV

StreamingMovie
Enter StreamingMovie

Payment Method
Enter Payment Method

Bank transfer (automatic) MonthlyCharge
Enter MonthlyCharge

TotalCharges
Enter TotalCharges

Clear Predict

Prediction The customer Remains stay

PDF IISPS_INT_2394_ED...pdf Show all

03:28 PM

7. ADVANTAGES AND DISADVANTAGES

Advantages:

- 📁 It is used in a variety of different industries and types of businesses.
- 📁 It is widely used for managing risks in the telecom customer company.

- ✎ It is composed using the HTML and Python for the web usage in real time.
- ✎ Random Forest gives the accurate result of the prediction upto 77% which is the algorithm we used for prediction.

Disadvantages:

- ✎ Gives only 77% accuracy for the customer churn status.
- ✎ The higher your customer churn rate, the lower the chances of growing the business.
- ✎ If they are losing customers at a high rate, as the cost of acquiring new customers is so high.

8. APPLICATIONS

- The term customer churn is used to describe the loss of existing customers
- Tracking customer churn is a key business metric for most companies.
- To have an idea of customer relationship cycle such as customer acquisition, increasing value of the customer and customer retention.
- It is widely used for managing risks in the telecom customer industry.
- So we use Machine Learning Algorithms to analyze the data and propose what company and churn lending companies need to achieve their needs.

9. CONCLUSION

Churn rate is a health indicator for subscription based companies. The ability to identify customers that aren't happy with provided solutions allows business to learn about product or pricing plan weak points, operation issues, as well as customer preferences and expectations to proactively reduce reasons for churn.

In this paper, the Random Forest algorithm is adopted to build a UI model for predicting customer churn default in the lending club and the results are compared with other six algorithms of logistic regression, KNN, naive bayes, decision tree and support vector machine. The experiment shows that the Random Forest algorithm performs

outstanding than the other six algorithms in the prediction of customer churn default and has strong ability of generalization. There is no definitive guide of which algorithms to use given any situation. What may work on some data sets may not necessarily work on others. Therefore, always evaluate methods using cross validation to get a reliable estimates.

10. FUTURE SCOPE

In future the Random Forest algorithm can be applied on other data sets available for telecom customer churn approvals to further investigate its accuracy. A rigorous analysis of other machine learning algorithms other than these six can also be done in future to investigate the power of machine learning algorithms for customer churn status prediction. In further study, we will try to conduct experiments on larger data sets or try to tune the model so as to achieve the state -of-art performance of the model and a great UI support system making it complete web application model.

11. BIBLIOGRAPHY

- Hands-On Machine Learning with Scikit-Learn and TensorFlow (2nd Edition) by Aurélien Géron
- The Hundred-Page Machine Learning Book by Andriy Burkov
- Building Machine Learning Powered Applications: Going from Idea to Product by Emmanuel Ameisen
- Grokking Deep Learning by Andrew W. Trask
- Deep Learning with Python by Francois Chollet
- Deep Learning by Ian Goodfellow, Yoshua Bengio, Aaron Courville
- Reinforcement Learning: An Introduction (2nd Edition) by Richard S. Sutton,

Andrew G. Barto

- Deep Reinforcement Learning Hands-On (2nd Edition) by Maxim Lapan
- TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers by Pete Warden & Daniel Situnayake
- Learning From Data by Yaser S. Abu-Mostafa, Malik Magdon-Ismael, Hsuan-Tien Lin.
- The Book of Why by Judea Pearl, Dana Mackenzie.
- Rebooting AI by Gary Marcus & Ernest Davis
- Machine Learning Yearning by Andrew Ng.
- An Introduction to Machine Learning Interpretability (2nd Edition) by Patrick Hall & Navdeep Gill
- Interpretable Machine Learning by Christoph Molnar.
- Neural Networks and Deep Learning by Michael Nielsen.
- Generative Deep Learning by David Foster

APPENDIX

HTML:

```
<!DOCTYPE html>
```

```
<html>
```

```
<!--From https://codepen.io/frytyler/pen/EGdtg-->
```

```
<head>
```

```
  <meta charset="UTF-8">
```

```
  <title>ML API</title>
```

```
<link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300'
rel='stylesheet' type='text/css'>
<link rel="stylesheet" href="../static/css/style.css">
```

```
<style>
    .login {
        top: 20%;
    }
</style>
</head>
```

```
<body>
<h1>Telecom Customer Churn Prediction</h1>
<div class="login">
```

```
<!-- Main Input For Receiving Query to our ML -->
<form action="{{ url_for('fun2')}}" method="post">
    <label for="SeniorCitizen"> SeniorCitizen</label>
    <input type="text" name="sc" placeholder="Enter SeniorCitizen" id="SeniorCitizen" required="required" />

    <label for="Partner">Partner</label>
```

```
<input type="text" name="pa" placeholder="Enter Partner" id="Partner"
required="required" />
```

```
<label for="tenure">tenure</label>
<input type="text" name="te" placeholder="Enter tenure" id="tenure"
required="required" />
```

```
<label for="PhoneService">PhoneService</label>
<input type="text" name="ps" placeholder="Enter PhoneService"
id="PhoneService" required="required" />
```

```
<label for="MultipleLines">MultipleLines</label>
<input type="text" name="ml" placeholder="Enter MultipleLines"
id="MultipleLines" required="required"/>
```

```
<label for="InternetService">InternetService</label>
<input type="text" name="ls" placeholder="Enter InternetService"
id="InternetService" required="required" />
```

```
<label for="TechSupport">TechSupport</label>
<input type="text" name="ts" placeholder="Enter TechSupport" id="TechSupport"
required="required" />
```

```
<label for="StreamingTV">StreamingTV</label>
<input type="text" name="st" placeholder="Enter StreamingTV" id="StreamingTV"
required="required" />
```

```
<label for="StreamingMovie">StreamingMovie</label>
<input type="text" name="sm" placeholder="Enter StreamingMovie"
```

```
id="StreamingMovie" required="required" />
```

```
<label for="Payment Method">Payment Method</label>
```

```
<input type="text" name="pm" placeholder="Enter Payment Method" id="Payment  
Method" required="required" />
```

```
<select name = "Payment Method">
```

```
<option value = "Bank transfer (automatic)">Bank transfer (automatic)</option>
```

```
<option value = "Credit card (automatic)">Credit card (automatic)</option>
```

```
<option value = "Electronic check">Electronic check</option>
```

```
<option value = "Mailed check">Mailed check</option>
```

```
</select>
```

```
<label for="MonthlyCharge">MonthlyCharge</label>
```

```
<input type="text" name="mc" placeholder="Enter MonthlyCharge"  
id="MonthlyCharge" required="required" />
```

```
<label for="TotalCharges">TotalCharges</label>
```

```
<input type="text" name="tc" placeholder="Enter TotalCharges" id="TotalCharges"  
required="required" />
```

```
<button type="reset" class="btn btn-primary btn-large">Clear</button>
```

```
<button type="submit" class="btn btn-primary btn-large">Predict</button>
```

```
<br>
```

```
<b>{{pred}}</b>
```

```
</form>
```

</div>

</body>

</html>

APP.PY:

```
from flask import Flask , render_template,request
import pickle
model = pickle.load(open(r'D:\srija\Flaskapp\Churn.pkl','rb'))
app = Flask(__name__)
@app.route('/') #when the browser is routed towards url execute below function
def y_pred():
    return render_template("index.html")
@app.route('/login',methods=["POST"])
def fun2():

    sc = request.form['sc']
    pa = request.form['pa']
    te = request.form['te']
    ps = request.form['ps']
    ml = request.form['ml']
    ls = request.form['ls']
    ts = request.form['ts']
    st = request.form['st']
    sm = request.form['sm']
```



```

pm = request.form['Payment Method']
if(pm=='Bank transfer (automatic)'):
    s1,s2,s3,s4=1,0,0,0
if(pm=='Credit card (automatic)'):
    s1,s2,s3,s4=0,1,0,0
if(pm=='Electronic check'):
    s1,s2,s3,s4=0,0,1,0
if(pm=='Mailed check'):
    s1,s2,s3,s4=0,0,0,1

mc = request.form['mc']
tc = request.form['tc']
data =
[[int(sc),int(pa),int(te),int(ps),int(ml),int(ls),int(ts),int(st),int(sm),int(s1),int(s2),int(s3),int(s
4),str(mc),str(tc)]]
pred=model.predict(data)
if(pred==[0]):
    ans = "The customer will exit"
else:
    ans = "The customer Remains stay"

return render_template("index.html", pred= 'Prediction {}'.format(ans))
if __name__ == '__main__':
    app.run(debug = True)

```