Comparative Analysis of Classification algorithms for Customer Retention in Telecom Industry

Sanjiv Kumar Upadhyay (19MCA0248)   Aniket Pandey (19MCA0202)   Shubham Pandey (19MCA0238)

Project Guide
Prof. DR. SHYNU P. G.

School of Information Technology & Engineering (SITE)
Vellore Institute of Technology (VIT)
Vellore, Tamil Nadu, India

Abstract

Companies which are based on Telecommunication is an industry which is making its presence felt period by period. And the more it grows, more links are getting connected in this chain. Because of business strategies and expansion of the companies willing to compete with other organizations for adding more clients to the network is an expensive process, but retention of the existing client in hand is 6 times less expensive when it is compared with the cost for making a new subscriber to the organization. With the usage of classification algorithms like Naïve Bayes, Decision Trees, Support Vector, Random forest and Artificial Neural Networks this

project will be successful to tell to tell that which among the used algorithm is most suitable for predicting the churning characteristics of the subscriber, which shall allow the company to do their rightful measures at an appropriate time by availing generous offers to existing subscriber and others techniques in telecommunication department so that the customer retention will be possible what's more, benefit is picked up.

Keywords: Customer Retention, Churn, Classification, Telecommunication Industries

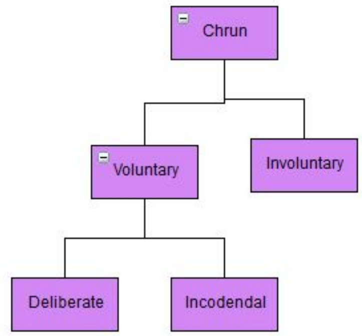
Introduction

Various Telecom Companies are collecting huge volumes of data about their current or potential subscribers, we can pass all these data to machine learning classification algorithms to get useful data from collected data over a period of time. Among others the most useful knowledge that Telecom Companies can extract is churn behavior of their customers. Churn customer means the person who is going to switch the telecom organization after comparing his current service profit with a potential switchable service provider. There are basically two types of churn.

Voluntary Churn:

It means the customer themselves took the decision to switch the company or stop using the services. They are partitioned into two categories.

1. **Deliberate churn**: These customer switching company because of the low price or good quality services are offered by competitors.



(Figure 1)

2. **Incidental churn**: It occurs when the subscriber changes its primary location, and the telecom organization doesn't have network coverage over that area.

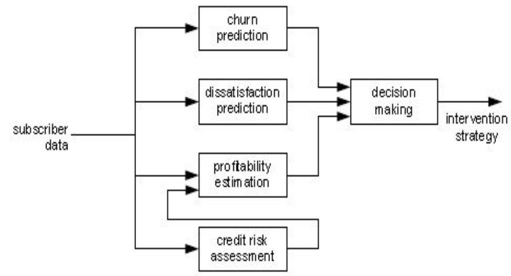
Literature survey

[1] The objective of the survey is to figure out which type of subscriber of a telecommunications company is most probable to go churn. In this survey the authors are using a decision tree algorithm by which they achieved a predictive accuracy of 82%. In the resulting data set, each row represents one customer of the company describing their telecommunication behavior for each of five months. Whether or not the client left the organization in the 6th month tells the binary classification label or target. When a learned classifier is accessible it very well may be applied each month on information from the present and recent month, to anticipate churn for the following month. The procedure model as set up with the

MiningMart customer is traded to a record dependent on XML linguistic structure, and sent to the focal "Case Base" that is kept up on the MiningMart site.

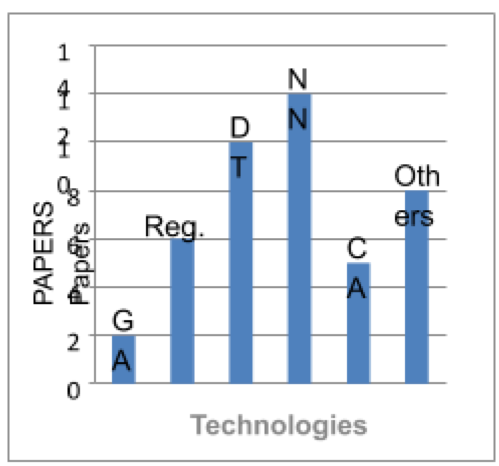
[2] The objective of the paper is to put lights on major operators to retain subscribers. Through observational assessment, this examination looks at different information mining technology to every endorser of a mobile administrator. Taiwan started showing its interest in Telecommunication by permitting licenses to nearly 6 mobile operators. The mediocre ratio of churn of a remote operator is about two percent every month. Explicitly with information mining systems finding the best model of predictive churn from data distribution center to forestall the client behavior. With adequate database size and quality, in light of the aftereffect of meetings with specialists, we remove some conceivable variables from the client database as a logical base of EDA to figure out which variable are helpful to separate among churners and non-churners.

[3] The objective of the paper is to apply mining algorithms like logit regression, decision trees, artificial neural networks and boosting algorithms. Trials depend on a database of almost 47000 U.S. local endorsers, and incorporate data about their use, charging, credit, application, and grumbling history. In light of the supporter anticipated agitate likelihood, we should conclude whether to offer the endorser some impetus to stay with the bearer, which will apparently decrease the every probable outcome of churn.



(Figure 3): The framework for churn prediction and profitability maximization

[4] The target of this survey is maintenance of the hand clients. This paper depends on the comparative study of a portion of the Data Mining calculations and discovers what one of those calculations is most appropriate for different kinds of organizations. The fast interest of better ideas in the market in each area is prompting a predominant endorser base for specialist co-op. The accompanying paper surveys the various methodologies utilized by specialists like Regression formula, Naïve Bayes, Decision Trees, Neural Networks, Cluster Analysis and Others for the communication sector as well other sectors which highly depend on customer participation. To solve this Problematic situation we have to identify churn before it becomes actual churns.



(Figure 4)

[5] Social organization Analysis has risen as a key worldview in present day human science, innovation, and data sciences, we study the development of churners in an administrator's system traversing over a time of four months. Our practical analysis explores the possibilities of a client to churn all out of a telecom service provider's network inspired by a pre-churned subscriber. This paper analyzes the correspondence examples of a huge number of cell phone clients, permitting us to examine the fundamental informal organization in a huge scope of alternative placements. This paper exhibits a basic, yet compelling, dissemination based methodology that misuses these impact to recognize a delicate part of churner's behaviour in the process.

[6] The target of the overview is cost decrease. Churn expectation models are ordinarily assessed utilizing factually based execution measures. A tale, benefit driven execution measure is created, by figuring the most extreme benefit that can be produced by including the ideal part of clients with the most noteworthy anticipated probabilities to be worn down in a maintenance crusade. At long last, an enormous gathering of classifiers is found to yield equivalent execution. Applying the most suitable benefit model and remembering the ideal portion of clients for a maintenance battle prompts generously various results.

[7] This paper has utilized information mining characterization strategies including Regression formula, Naïve Bayes, Decision Trees, Neural Networks, SVM in order to have a look about their exhibitions. Paper is utilizing the information of an Iranian versatile organization, not exclusively were these strategies experienced and contrasted with one. Breaking down the procedures' conduct and coming to get information about their claims to specialties, they proposed a cross breed approach which made more exact upgrades to the estimation of the assessments measurements. The proposed nearest suitable technique, results indicated that above ninety five % exactness for Recall and Precision is properly feasible. Every classification was calculated in tabular form and the proposed hybrid solution seems that gaining a remarkably good Precision means a low Recall measure and vice versa.

[8] In this paper we came to think about existing fraud recognition strategies that are investigated and another fraud recognition technique is suggested. The selection of the three classification algorithmic model and one hybrid meta-learning technique is justified for the new method. By using this algorithms to process the sampled information partitions, and by methods for a direct cost model to assess the classifiers with, the best among those classifiers ought to likely be picked for sending inside an all-around shaped association. Stacking bagging additionally beats the normal method utilized in industry. In this way, this paper looks at the currently available fraud detection technique against C4.5.

Dataset Attributes

Tenure	Value Type: Integer ,Value: Greater than 0.0
Phone Service	Value Type: Boolean ,Value: Yes or No
Multiple Lines	Value Type: String , Value : Yes, No or No, Phone Service
Internet Service	Value Type: String Value: DSL, Fibre optic, No
Online Security	Value Type: String, Value : Yes, No or No, Internet Service
Online Backup	Value Type: String, Value : Yes, No or No, Internet Service
Device Protection	Value Type: String, Value : Yes, No or No, Internet Service
Tech Support	Value Type: String, Value : Yes, No or No, Internet Service
Streaming TV	Value Type: String, Value : Yes, No or No, Internet Service
Streaming Movies	Value Type: String, Value : Yes, No or No, Internet Service
Contract	Value Type: String, Value: One Year , Month-to-month, Two year
Paperless Billing	Value Type: Boolean, Value : Yes or No
Payment Method	Value Type: String Value: Electronic check, Mailed check, Bank transfer, Credit card

Attribute	**Purpose**
Customer ID	Type: String Value: Unique
gender	Type: String Value : Male or Female
Senior Citizen	Type: Integer Value : 1 or 0
Partner	Type: Boolean Value : Yes or No
Dependents	Type: Boolean Value : Yes or No

Data Set

The dataset we have used for this paper is provided by Telecom Subscriber Churn. Every single row represents a customer and each vertical section shows customer attributes. There are a total of 7043 rows and 21 columns. For our practical analysis we will be using 20 columns only. The “churn” column is a label attribute.

The dataset include information about following:

- Subscribers who left last month are represented by the column “churn”.

- List of attributes are present for service oriented information.
- Customer account related information is also provided.

Monthly Charges	Type: Decimal Value :Greater than 0.0
Total Charges	Type: Decimal Value :Greater than Zero
Churn	Type: Boolean Value: Yes or No

Experiment Result

The experiment is performed on JupyterLab version 1.1.4. Python3 is used for programming the structure of following algorithms.

Data Preprocessing

Before proceeding toward the data set we need to process the input values for absent values. The provided dataset does not have any row with all the missing values. No two rows are the same. But we take care of null values. During processing of data we replace null values with the mean of the particular column. The reason behind this is to avoid outliers and spreading of data.

The dataset contains categorical and continuous features. All the categorical features are ordinal in nature. We have replaced Yes with 1 and No with 0. Other

- Demographic information of customers is also included in the dataset.

features may have more value relying on the number of special data present in the feature

Data Analysis

Before starting to create models we need to analyze our data so that we can create our model accordingly. Analyzing helps in determining features who are affecting the target most. Python provides a set of tools which can be very helpful in plotting graphs and charts. One of the tools is matplot. For this experiment we have used matplot library for plotting graphs so that we can visualize data more accurately.

Categorical Features

The following are the categorical features present in the telco customer churn dataset.

- gender
- Senior Citizen
- Partner
- Dependents
- Phone Service
- Multiple Lines
- Internet Service
- Online Security
- Online Backup
- Device Protection
- Tech Support
- Streaming TV

- Streaming Movies
- Contract
- Paperless Billing
- Payment Method

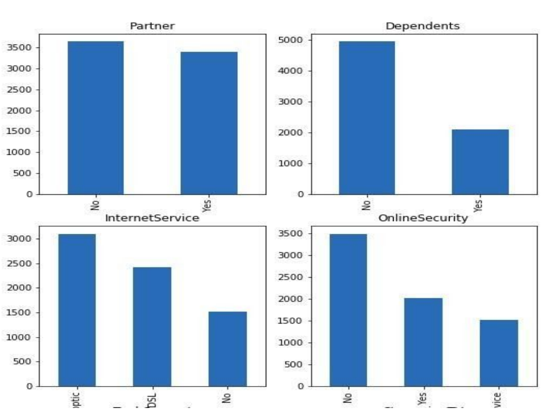
Data Summary

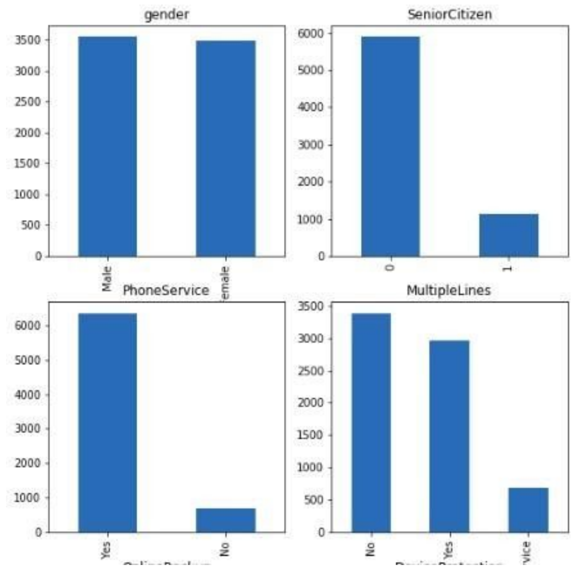
The summary presents a list of properties with corresponding values which can help understanding the data. The summary enlightens us in understanding the spread of information. The thickness of information can be found by the quartiles values.

	gender	SeniorCitizen	Partner
count	7010.000000	7010.000000	7010.000000
mean	0.504280	0.162767	0.484023
std	0.500017	0.369180	0.499780
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	1.000000	0.000000	0.000000
75%	1.000000	0.000000	1.000000
max	1.000000	1.000000	1.000000

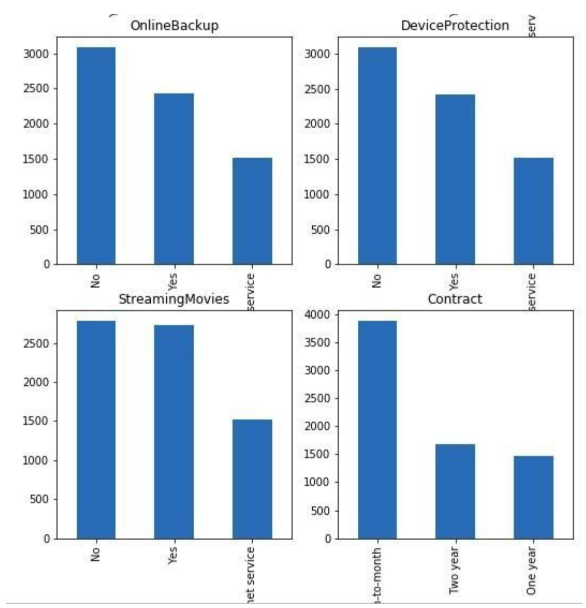
(Figure 5)

Data Visualization

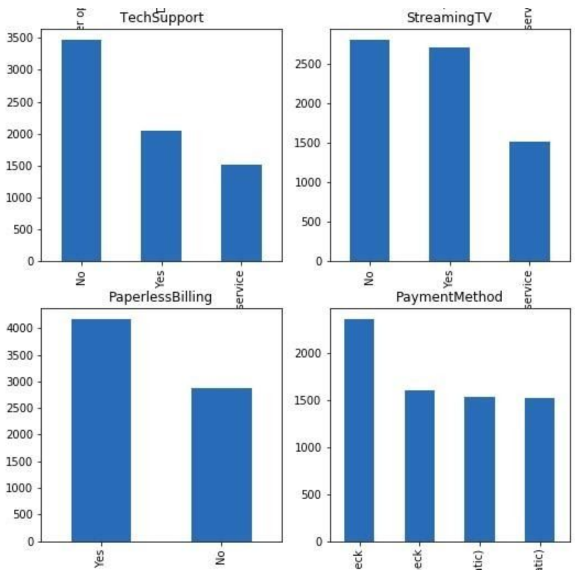




(Figure 7)

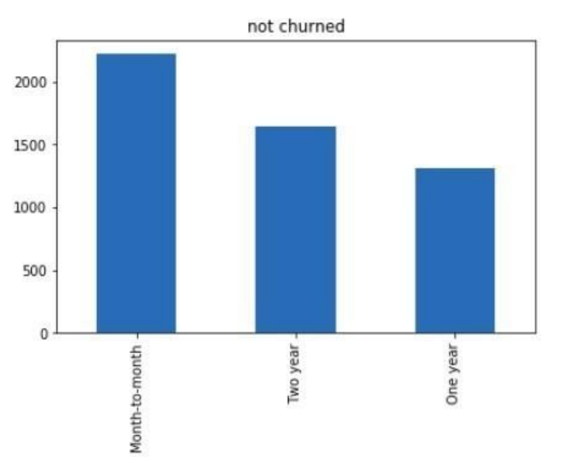


(Figure 8)

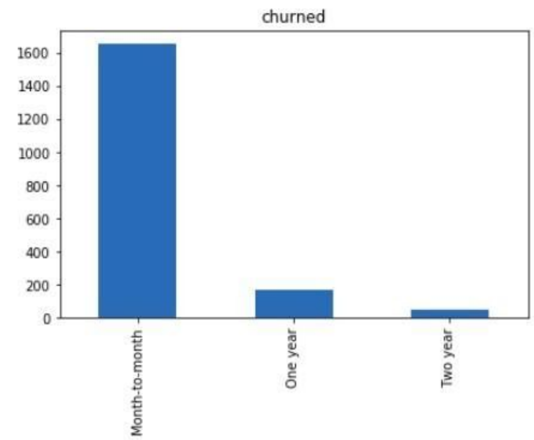


(Figure 9)

The Contract has three different values “Month to month”, “Two year” and “One year”. If you try to find the relationship between “Contract” attribute and the target attribute then we can make a conclusion that those customers are more likely to churn which are under Month to month contracts.



(Figure 10)



(Figure 11)

Continuous Features

The following are the continuous features which are present in the telco customer churn dataset.

- Total Charges
- Tenure
- Monthly Charges

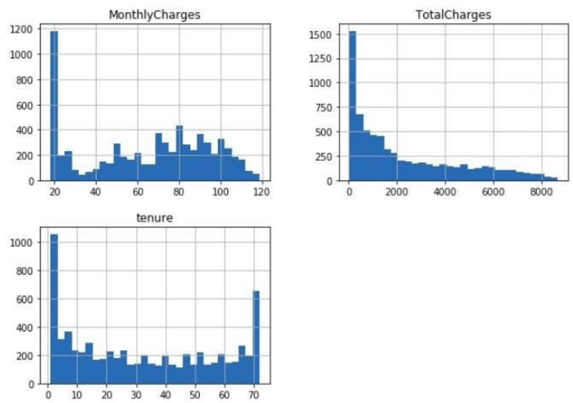
Data Summary

	TotalCharges	tenure	MonthlyCharges
count	7032.000000	7032.000000	7032.000000
mean	2283.300441	32.421786	64.798208
std	2266.771362	24.545260	30.085974
min	18.800000	1.000000	18.250000
25%	401.450000	9.000000	35.587500
50%	1397.475000	29.000000	70.350000
75%	3794.737500	55.000000	89.862500
max	8684.800000	72.000000	118.750000

(Figure 12)

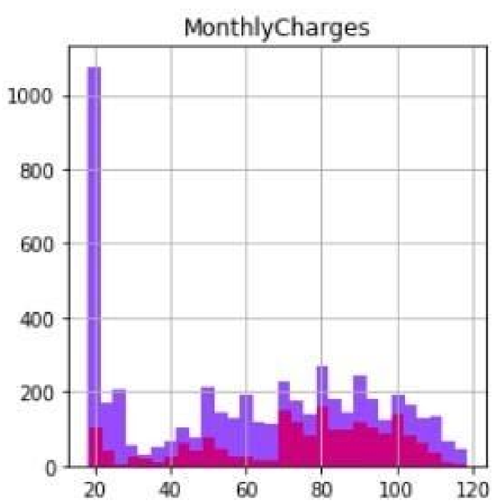
Data Visualization

The below histogram has been initialized to create 30 bins from minimum to maximum value of the attribute.

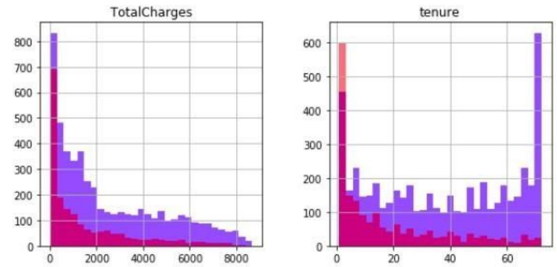


(Figure 13)

The below histogram provides us the relationship between the continuous attribute and the target attribute. After studying the graph we can see that customers who have high tenure and high total charges are more likely not to churn.



(Figure 14)



(Figure 15)

Data Modeling

This method is a known phenomenon which is used to train a model with training data. The purpose of the training model is able to predict the output of by providing a set of attributes.

Decision Tree

It is a one of the popular studies as a supervised learning algorithm. This algorithm attempts to take care of a difficult utilizing tree portrayal. Individual node of the portrayed tree is the attribute and each leaf node is a class label. The popular attribute selection measure are

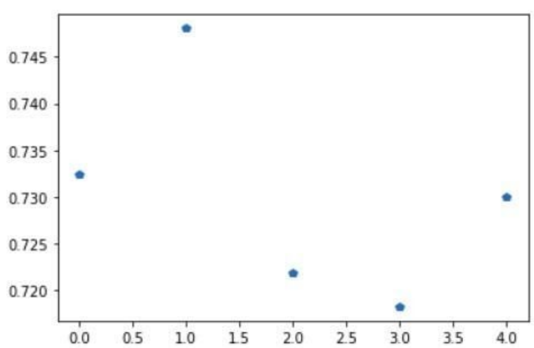
- Gini index
- Information Gain

Decision Tree can lead to overfitting. By applying the provided dataset after processing on this algorithm, we were able to get a model with the accuracy of 74%. A detailed report is provided below.

	precision	recall	f1-score	support
0.0	0.85	0.79	0.82	798
1.0	0.47	0.55	0.51	259
accuracy			0.74	1057
macro avg	0.66	0.67	0.66	1057
weighted avg	0.75	0.74	0.74	1057

(Figure 16)

Instead of having high accuracy the precision is quite low for finding churn customers. Cross validate is a technique in which a model is trained a number of times with different sets of training data chosen randomly.



(Figure 17)

Naive Bayes

Naive Bayes comes under the category of supervised learning algorithms. This algorithm uses probability to train the model. The base of this algorithm comes from the mathematical Bayes Theorem.

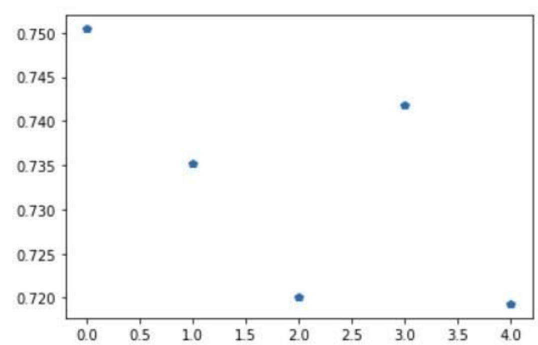
$$P(A|B) = P(B|A)P(A) / P(B)$$

Here the symbolism of A and B are two individual independent events. B is the evidence or B already happened and A is the hypothesis. A variant of this algorithm is used when the input features contain continuous values. Gaussian Naive Bayes is used for continuous values. After training the model with data, 74% accuracy is achieved.

	precision	recall	f1-score	support
0.0	0.93	0.71	0.81	798
1.0	0.48	0.83	0.61	259
accuracy			0.74	1057
macro avg	0.71	0.77	0.71	1057
weighted avg	0.82	0.74	0.76	1057

(Figure 18)

Decision trees and Naive Bayes both share the same accuracy rate but when we look at the final precise report we can find that Naive Bayes is more precise than Decision Tree.



(Figure 19)

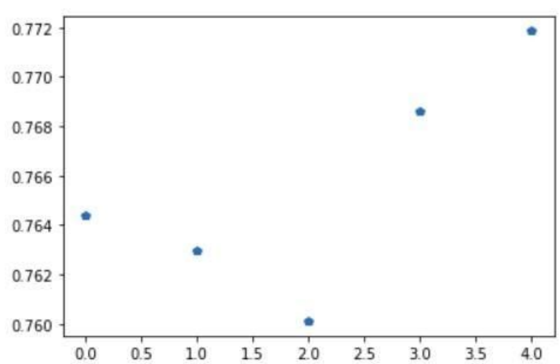
Support Vector Machine

SVM is also like others a supervised learning algorithm. This algorithm could be used for both regression as well as classification. But mainly this is being used for classification. The algorithm tries to find an optimal hyperplane between the sets of points which can help in classification. This line helps in segregating n-dimensional space. Training this model with "rbf" as kernel function and degree as 3 we are able to achieve an accuracy of 78%.

	precision	recall	f1-score	support
0.0	0.80	0.95	0.87	798
1.0	0.64	0.25	0.36	259
accuracy			0.78	1057
macro avg	0.72	0.60	0.62	1057
weighted avg	0.76	0.78	0.75	1057

(Figure 20)

Apart from accuracy this model also proves better in precision for finding churn customers. The precision in finding churn customers in the training dataset is 64% which is higher than 48% and 47% of Naive Bayes and Decision Tree respectively.



(Figure 21)

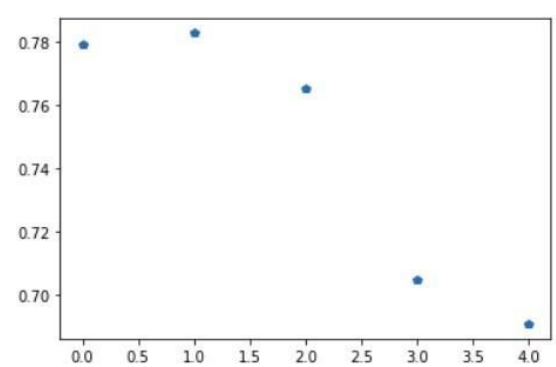
Neural Network

NN is a classification algorithm which is a group of neurons. These neurons shall have several unidentified hidden layers. Containing more than one hidden layer is also called multi-layer neurons. In our experiment also we have used MLP network for training the model and we are able to achieve an accuracy of 81%. We run the model for 200 epocs and set the model to terminate when loss is not decreasing for several epochs.

	precision	recall	f1-score	support
0.0	0.86	0.90	0.88	798
1.0	0.63	0.53	0.58	259
accuracy			0.81	1057
macro avg	0.74	0.72	0.73	1057
weighted avg	0.80	0.81	0.80	1057

(Figure 22)

The overall accuracy is more than the earlier three models. But precision is finding non churn is lesser than Naive Bayes. This can be solved by training the model with more data.



(Figure 23)

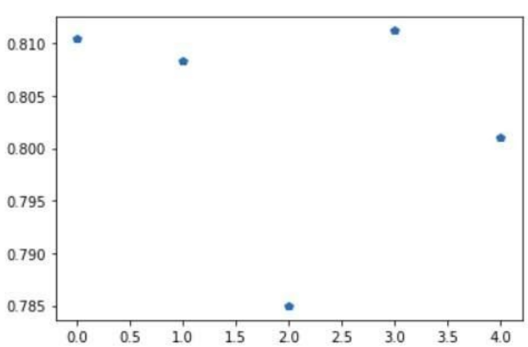
Random Forest

Random Forest is an application of an ensemble model. **Ensembling** is a technique in which more than one model is created and trained. The input feature is provided to all the available models and the output is collected from them. The class which is having more votes is considered as the output. In our experiment we have created 300 estimators. The model is able to achieve 82% accuracy.

	precision	recall	f1-score	support
0.0	0.87	0.90	0.88	798
1.0	0.65	0.57	0.61	259
accuracy			0.82	1057
macro avg	0.76	0.74	0.75	1057
weighted avg	0.81	0.82	0.82	1057

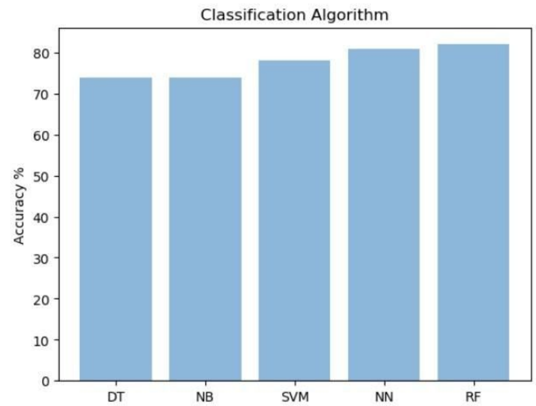
(Figure 24)

The model is able to achieve the highest accuracy among all. But the precision for non-churn customers is lower than Naive Bayes.



(Figure 25)

The below graph shows the accuracy of all the classification algorithms which are discussed above.



(Figure 28)

DT: Decision Tree

NB: Naive Bayes

SVM: Support Vector Machine

NN: Neural Network

RF: Random Forest

Conclusion

After implementing all five classification algorithms we can conclude that Random Forest is the most accurate classification algorithm for Telco Customer Churn dataset. Few improvements can be done to increase the accuracy of the models. The training data can be increased to train the model more accurately. Overfitting should be avoided. SVM in higher dimensions may perform better.

References

[1] Euler, T., 2005, October. Churn prediction in telecommunications using miningmart. In *Proceedings of the workshop on data mining and business (DMBiz)*.

[2] Hung, S.Y., Yen, D.C. and Wang, H.Y., 2006. Applying data mining to telecom churn management. *Expert Systems with Applications*, *31*(3), pp.515-524.

[3] Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H., 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on neural networks*, *11*(3), pp.690-696.

[4] Kamalraj, N. and Malathi, A., 2013. A survey on churn prediction techniques in the communication sector. *International Journal of Computer Applications*, *64*(5), pp.39-42.

[5] Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A.A. and Joshi, A., 2008, March. Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology* (pp. 668-677).

[6] Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B., 2012. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), pp.211-229.

[7] Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M. and Abbasi, U., 2014. Improved churn prediction in the telecommunication industry using data mining techniques. *Applied Soft Computing*, *24*, pp.994-1012.

[8] Phua, C., Alahakoon, D. and Lee, V., 2004. Minority report in fraud detection: classification of skewed data. *Acm sigkdd explorations newsletter*, *6*(1), pp.50-59.

[9] Coussement, K. and Van den Poel, D., 2008. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications*, *34*(1), pp.313-327.

[10] Au, W.H., Chan, K.C. and Yao, X., 2003. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE transactions on evolutionary computation*, *7*(6), pp.532-545.