International Journal of Future Generation Communication and Networking
Vol. 14, No. 1, (2021), pp. 1482-1495

Customer Churn Prediction Using Machine Learning

D. Deepika[1], Nihal Chandra[2]
Assistant Professor, Department of CSE, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, INDIA[1]
Student, Mahatma Gandhi Institute of Technology,Hyderabad,Telangana,INDIA[2]
deshmukhdeepika@gmail.com[1]

ABSTRACT

Customers are the most important assets in any industry since they are considered as the main profit source. Companies are working hard to survive in today's competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues: (1) acquiring new customers, (2) upselling the existing customers, and (3) increasing the retention period of customers. However, comparison of these strategies has shown 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.Customer churn is a term that refers to the loss of a client or customer—that is, when a customer ceases to interact with a company or business. Similarly, the churn rate is the rate at which customers or clients are leaving a company within a specific period of time. Customer churn is 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 that could churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn.

A churn prediction model is developed in this project which can assist companies to predict customers who are most likely to churn. It uses machine learning techniquessuch as Logistic Regression, Decision Trees, K-Nearest Neighbors and Support Vector Machine algorithmsto identify the primary determinants of customer churn along with the algorithm fit for such predictions.The dataset contains demographic details of customers, their total charges and they type of service they receive from the company. Itcomprises of churn data of over 7000 customersspread over 21 attributes obtained from Kaggle. Further on this investigation, the usage of the above mentioned algorithms is described for predicting customer churn.

To conclude, the results of the algorithmsfor predicting customer churn are outlined in the form of accuracy, recall score, precision, f1scoreand kappa metrics using interactive graphs. The results show that month-to-month contracts and the tenure of customers are most crucial attributes in predicting customer churn and an accuracy of 80.2% for Logistic Regression was the best.

INTRODUCTION

Customer retention is one of the primary growth pillars for products with a subscription-based business model. Competition is tough in markets where customers are free to choose from plenty of providers even within one product category. Several bad experiences – or even one – and a customer may quit. And if droves of unsatisfied customers churn at a clip, both material losses and damage to reputation would be enormous.

Telephone service companies, Internet service providers, pay TV companies, insurance firms, and alarm monitoring services, often use customer attrition analysis and customer attrition rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Companies from these sectors often have customer service branches which attempt to win back defecting clients, because recovered long-term customers can be worth much more to a company than newly recruited clients.

Companies usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another company or service provider, involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care facility, death, or the relocation to a distant location. Analysts tend to concentrate on voluntary churn, because it typically occurs due to factors of

the company-customer relationship which companies control, such as how billing interactions are handled or how after-sales help is provided.

Churn rate is a health indicator for businesses whose customers are subscribers and paying for services on a recurring basis. Customers (of subscription-driven businesses) opt for a product or a service for a particular period, which can be rather short – say, a month. Thus, a customer stays open for more interesting or advantageous offers. Plus, each time their current commitment ends, customers have a chance to reconsider and choose not to continue with the company. Some natural churn is inevitable, and the figure differs from industry to industry. But having a higher churn figure than that is a definite sign that a business is doing something wrong.Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base thatis most vulnerable to churn.

Motivation

There are many things companies may do wrong, from complicated onboarding when customers aren't given easy-to-understand information about product usage and its capabilities to poor communication, e.g. the lack of feedback or delayed answers to queries. Evenlongtime clients may feel unappreciated because they don't get as many bonuses as the new ones.In general, it's the overall customer experience that defines brand perception and influences how customers recognize value for money of products or services they use.Hence, this project uses churn prediction models to predict customer churn so that the above shortcomings can be overcome for any potential client by assessing their propensity of risk to churn.

Problem Definition

Managing customer churn is one major challenge facing companies, especially those that offer subscription-based services. Customer churn (or customer attrition) is basically the loss of customers, and it is caused by a change in taste, lack of proper customer relationship strategy, change of residence and several other reasons.If businesses can effectively predict customer attrition, they can segment those customers that are highly likely to churn and provide better services to them. Hence, a churn prediction model is developed in this project that uses machine learning techniques such as Logistic Regression, Decision Trees, K-Nearest Neighbors and Support Vector Machine algorithms to assist companies in predicting customers who are most likely to churn. In this way, they can achieve a high customer retention rate and maximize their revenue.

Existing System

Many approaches were applied to predict churn in telecom companies. Most of these approaches have used machine learning and data mining. For example, a churn prediction model for prepaid customers was developed in telecom using fuzzy classifiers, Neural Networks, SVM Classifier, Ada Boost & RF techniques, which were compared with a fuzzy nearest-neighbor classifier to predict an accurate set of churners on a real-time dataset of prepaid telecom customers from south Asia. Another model utilized a CRM frame work using neural network and data mining for the prediction of customer behavior in banking. An algorithm was also developed based on click stream data of a website to extract information and tested the predictive power of the model based on data such as number of clicks, repeated visits, repetitive purchases, etc. Nonetheless, these models raised a few concerns which are to be addressed. The main drawbacks of existing systems include:

Most of them were suited only for applying suitable model and taking inference from predictions.None of them focused on the attributes crucial towards customer churn.Focus was more towards comparison rather than attributes determination.

Proposed System

The basic model for predicting future customer churn is data from the past. We look at data from customers that already have churned (response) and their characteristics / behaviour (predictors) before the churn happened. The dataset contains demographic details of customers, their total charges and they type of service they receive from the company. It comprises of churn data of over 7000 customers spread over 21 attributes obtained from Kaggle[2].By fitting statistical models that relate the predictors to the response, we will try to predict the response for existing customers. This method belongs to the supervised learning category.

Requirements Specification

Software Requirements

Language : Python 3.6

Operatingsystem : Windows / Linux / macOS

Tools:

- Anaconda Navigator
- Jupyter Notebook
    - Numpy
    - Pandas
    - Matplotlib
    - Plotly

Hardware Requirements

RAM – 4GB minimum

Python

Python is a high-level, interpreted, interactive and object-oriented scripting languagecreated by Guido Rossum in 1989. Python is designed to be highly readable. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.It is ideally designed for rapid prototyping of complex applications.

It has interfaces to many OS system calls and libraries and is extensible to C or C++. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python programming is widely used in Artificial Intelligence, Natural Language Generation, Neural Networks and other advanced fields of Computer Science.

Pandas

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis / manipulation tool available in any language.

Pandas is well suited for many different kinds of data:

Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet Ordered and unordered (not necessarily fixed-frequency) time series data. Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels

Any other form of observational / statistical data sets. The data actually need not be labeled at all to be placed into a pandas data structure The two primary data structures of pandas, Series (1-dimensional) and DataFrame (2-dimensional), handle the vast majority of typical use cases in finance, statistics, social science, and many areas of engineering. For R users, DataFrame provides everything that R's data.frame provides and much more. pandas is built on top of NumPy and is intended to integrate well within a scientific computing environment with many other 3rd party libraries.

Matplotlib

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.

One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc. For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Plotly

Plotly is an interactive, open-source, and browser-based graphing library for Python. Built on top of plotly.js, plotly.py is a high-level, declarative charting library. plotly.js ships with over 30 chart types, including scientific charts, 3D graphs, statistical charts, SVG maps, financial charts, and more. Plotly has got some amazing features that make it better than other graphing libraries:

It is interactive by default

Charts are not saved as images but serialized as JSON, making them open to be read with R, MATLAB, Julia and others easily Exports vector for print/publication Easy to manipulate/embed on web

Anaconda Navigator

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. It is available for Windows, macOS, and Linux.

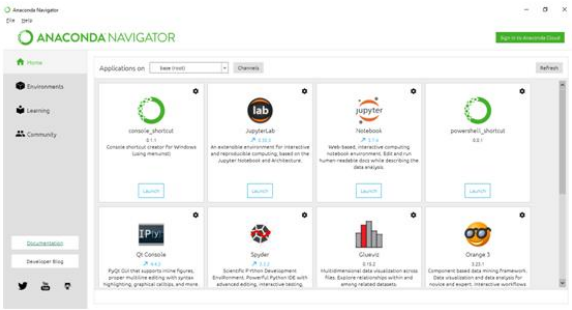

Figure 1 Anaconda Navigator window on Windows operating system

Executing the code with Navigator

The simplest way is with Spyder. From the Navigator Home tab, click Spyder, and write and execute your code.

You can also use Jupyter Notebooks the same way. Jupyter Notebooks are increasingly popular systems that combine your code, descriptive text, output, images, and interactive interfaces into a single notebook file that is edited, viewed, and used in a web browser.

The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

The Jupyter notebook combines two components:

A web application: A browser-based tool for interactive authoring of documents which combine explanatory text, mathematics, computations and their rich media output.

Notebook documents: A representation of all content visible in the web application, including inputs and outputs of the computations, explanatory text, mathematics, images, and rich media representations of objects.

Main features of the web application

In-browser editing for code, with automatic syntax highlighting, indentation, and tab completion/introspection. The ability to execute code from the browser, with the results of computations attached to the code which generated them.

Displaying the result of computation using rich media representations, such as HTML, LaTeX, PNG, SVG, etc. For example, publication-quality figures rendered by the matplotlib library, can be included inline.

In-browser editing for rich text using the Markdown markup language, which can provide commentary for the code, is not limited to plain text.

The ability to easily include mathematical notation within markdown cells using LaTeX, and rendered natively by MathJax.

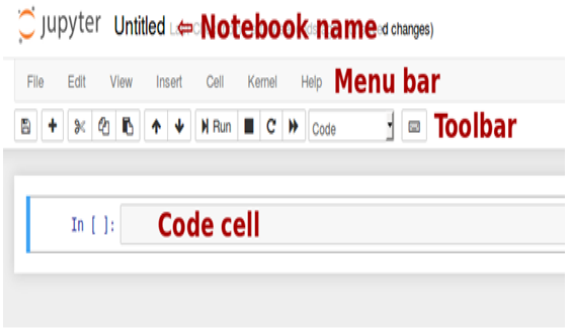

Figure 2 Jupyter Notebook Interface

LITERATURESURVEY

Potential research work carried out on various techniques for churn prediction in different areas such as telecom, e-commerce, and banking etc. has been discussed inthefollowingparagraphs. Various researchers have employed different mechanisms for predicting customer churn and to find out the most useful features used in the prediction.

Muhammad Azeem, Muhammad Usman and A. C. M. Fongpublished a paper titled"Achurnpredictionmodelfor prepaid customers in telecom using fuzzy classifiers". In this paper, a fuzzy based churn prediction model has been proposed and validated using a real data from a telecom company in South Asia. A number of predominant classifiers namely, Neural Networks, Linear Regression, C4.5, Support Vector Machines, AdaBoost, Gradient Boosting and Random Forest have been compared with fuzzy classifiers to highlight the superiority of fuzzy classifiers in predicting the accurate set of churners. Parameters such as TP rate & AUC were considered and enhanced using the model.

J. Vijaya and E. Sivasankar published a paper titled"An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing".It employs particle swarm optimization (PSO) and proposes three variants of PSO for churn prediction namely PSO incorporated with feature selection as its pre-processing mechanism, PSO embedded with simulated annealing and finally PSO with a combination of both feature selection and simulated annealing. The proposed classifiers were compared with decision tree, naive bayes, K-nearest neighbor, support vector machine, random forest and three hybrid models to analyze their predictability levels and performance aspects. Experiments reveal that the performance of meta-heuristics was more efficient and they also exhibited better predictability levels.

Matrin Fridrich published a paper titled "Hyper parameter Optimization of Artificial Neural Network in Customer Churn Prediction" using Genetic Algorithm sine-commerce. The prediction model is developed to identify customers at risk of defection. The proposed model leads to improved customer churn prediction ability on the basis of parameters such as TP rate, FP rate, and accuracy. The analysis is carried out on labeled ecommerce retail dataset describing 10,000 most valuable customers with the highest CLV (Customer Lifetime

Value). To obtain the best performing ANN (Artificial Neural Network) classification model, the proposed hyperparameter search space is explored with genetic algorithm to find suitable parameter settings. Explored part of hyperparameter search space is analyzed with conditional inference tree structure addressing underlying fundamental context of given optimization which results in identification of critical factors leading to well performing ANN classification model.

Gordini and Veglio published a paper titled "Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2Be-commerce industry". Parameters such as recentness, frequency, length, product category, failure, monetary, age, profession, gender, request status etc. Were taken for performance comparison. The prediction power of the proposed method was found to be better as compared to Linear Regression, Neural Networks & SVM especially for noisy, imbalance & nonlinear data. Thus, their findings confirm that the data-driven approach to churn prediction and the development of retention strategiesoutperforms commonly used managerial heuristics in B2B e-commerce industry.

Femina Bahari and Sudheep Elayidom published a paper titled"An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour" in banking. The UCI dataset containing direct bank marketing campaigns of Portuguese bank was taken. The model is used to predict the behaviour of customers to enhance the decision-making processes for retaining valued customers. Two classification models, Naïve Bayes and Neural Networks are studied and it was concluded that Neural Network was better than Naïve Bayes algorithm for accuracy & specificity while Naïve Bayes was better than Neural Network algorithm for sensitivity, TPR, FPR, and ROC area. Neural network classified 4007/514 & Naive Bayes classified 3977/544 instances correctly/incorrectly.

CUSTOMER CHURN PREDICTION METHODOLOGY

The basic layer for predicting future customer churn is data from the past. We look at data from customers that already have churned (response) and their characteristics / behaviour (predictors) before the churn happened. By fitting a statistical model that relates the predictors to the response, the response for existing customers is predicted. The overall scope of work to forecast customer attrition may look like the following:

Understanding a problem and final goal

- Data collection
- Data preparation and preprocessing
- Modeling and testing
- Model deployment and monitoring
- Understanding a problem and a final goal

It's important to understand what insights one needs to get from the analysis. In short, you must decide what question to ask and consequently what type of machine learning problem to solve: classification or regression.

Classification

The goal of classification is to determine to which class or category a data point (that is, customer) belongs to. For classification problems, historical data is used with predefined target variables, that is, labels (churner/non-churner) – answers that need to be predicted – to train an algorithm. With classification, businesses can answer the following questions:

Will this customer churn or not?

Will a customer renew their subscription?

Will a user downgrade a pricing plan?

Are there any signs of unusual customer behavior?

Regression

Customer churn prediction can be also formulated as a regression task. Regression analysis is a statistical technique to estimate the relationship between a target variable and other data values that influence the target variable, expressed in continuous values. The result of regression is always some number, while classification always suggests a category. In addition, regression analysis allows for estimating how many different variables in data influence a target variable. With regression, businesses can forecast in what period of time, a specific customer is likely to churn or receive some probability estimate of churn per customer.

Data collection

Once kinds of insights to look for are identified, the data sources necessary for further predictive modeling can be decided. The dataset used for this projectcontains demographic details of customers, their total charges and they type of service they receive from the company. It comprises of churn data of over 7000 customers spread over 21 attributes (described in Table 3.1) obtained from the Kaggle website.(as shown in Figure 3.1). It can be used to analyze all relevant customer data and develop focused customer retention programs.

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No

Figure 3 A snapshot of the dataset being used.

In the given Figure 3 each row represents a customer, and each column contains customer's attributes described on the column Metadata.

The data set includes information about:

Customers who left within the last month – the column is called Churn Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies

Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges

Demographic info about customers – gender, age range, and if they have partners and dependents.

Modeling and Testing

The main goal of this project stage is to develop a churn prediction model. Specialists usually train numerous models, tune, evaluate, and test them (as shown in the figure below) to define the one that detects potential churners with the desired level of accuracy on training data.The following supervised machine learning models have been used for predicting customer churn:

Logistic Regression

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable. Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled "0" and "1".

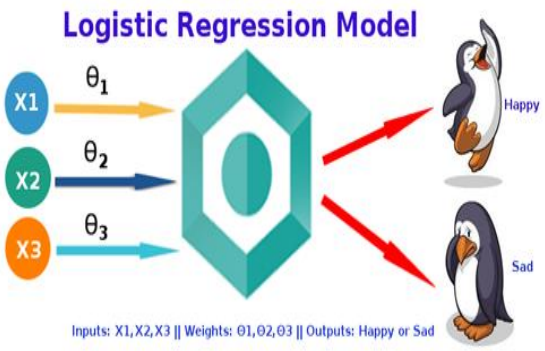

Figure 4 The binary logistic regression model basically gives you two possible values – 0/1, happy/sad and churn/not churn.

Logistic Function

Logistic regression is named for the function used at the core of the method, the logistic function.

The logistic function, also called the sigmoid function is an S-shaped curve (as shown in Figure) that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

Where e is the base of the natural logarithms and t value is the actual numerical value that you want to transform.

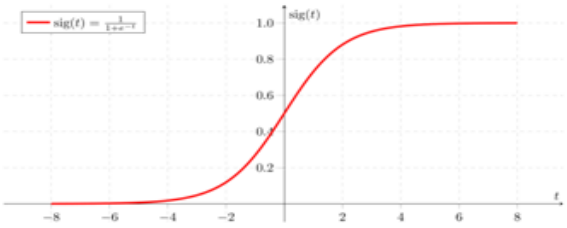

Figure 6 Logistic Function

Decision Trees

Decision tree learning is one of the predictive modeling approaches that uses a decision tree (as a predictive model) to go from observations about an item i.e. attribute (represented in the branches) to conclusions about the item's target value i.e. churn or not (represented in the leaves). Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.This algorithm splits a data sample into two or more homogeneous sets based on the most significant differentiator in input variables to make a prediction. With each split, a part of a tree is being generated. As a result, a tree with decision nodes and leaf nodes (which are decisions or classifications) is developed. A tree starts from a root node – the best predictor.

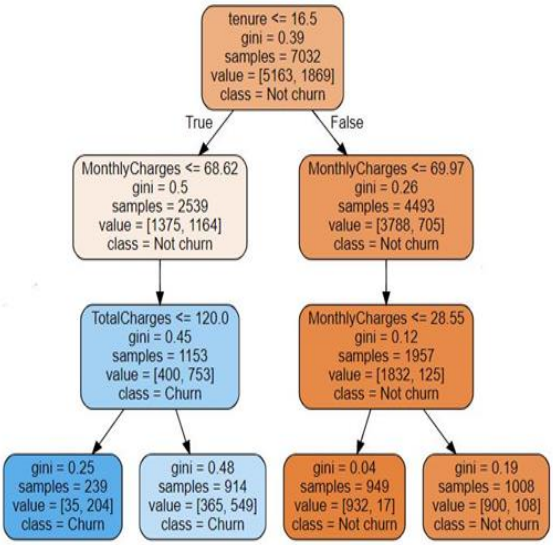

Figure 5 Basic structure of a decision tree

Prediction results of decision trees can be easily interpreted and visualized. Even people without an analytical or data science background can understand how a certain output appeared. Compared to other algorithms, decision trees require less data preparation, which is also an advantage. However, they may be unstable if any small changes were made in data. In other words, variations in data may lead to radically different trees being generated.

UMLDiagrams

Data Flow Diagram

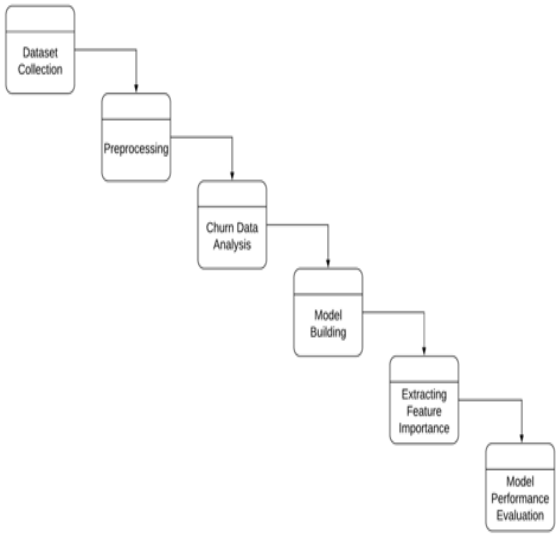

Figure 7 Data Flow Diagram for Churn Prediction

The data flow diagram shownin Figure 3.8 shows the flow of data right from acquiring the dataset to model building, extracting feature importance and comparison of model performances.

Sequence Diagram

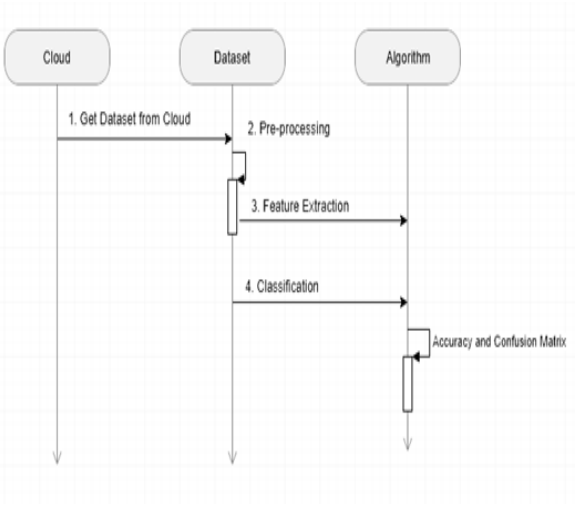

Figure 8 Sequence Diagram for Customer Churn Prediction

The sequence diagram shown in Figure 3.9shows the sequence of executing of various processes wiz. acquiring dataset, preprocessing, feature extraction, predicting results and accuracy determination.

Activity Diagram

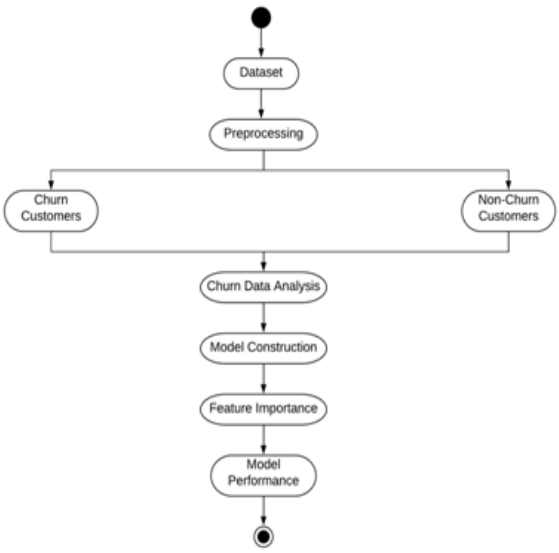

Figure 10 Activity diagram for implementation of Churn Prediction algorithm

The activity diagram shown in Figure 3.10 displays the flow of execution right from dataset gathering to training the algorithm to feature evaluation and analysis of results.

Collaboration Diagram

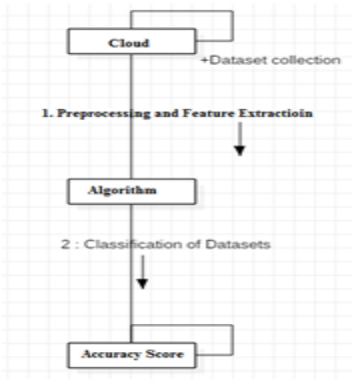

Figure 11 Collaboration diagram for Customer Churn Prediction

The collaboration diagram shown in Figure illustrates the relationships and interactions among the data and the algorithm of the prediction model.

TESTING ANDRESULTS

Testing is a crucial phase that determines the quality of models used as well as the importances of all the features under consideration. The algorithms used in this project have been rigorously tested based on various factors including accuracy, recall, precision, f1 score and kappa statistic. Accuracy - It measures how many observations, both positive and negative, were correctly classified.

$$Accuracy = \frac{(tp + tn)}{(tp + fp + fn + tn)} \quad \ldots(1)$$

For Logistic Regression,

$$Accuracy = \frac{259 + 1152}{259 + 231 + 116 + 1152} = 0.802 = 80.2\%$$

For KNN Classifier,

$$Accuracy = \frac{351 + 878}{351 + 139 + 390 + 878} = 0.699 = 69.9\%$$

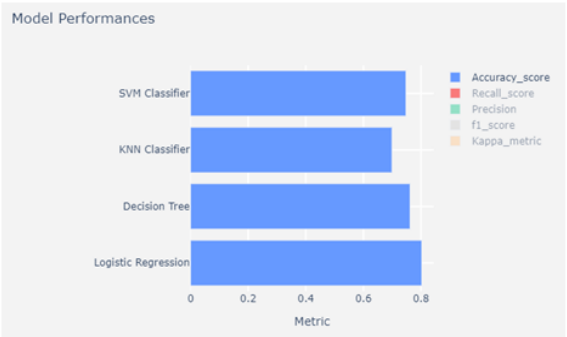

Figure 12. From the above figure, it is clear that Logistic Regression has the highest accuracy of 0.802 while the KNN classifier performed the worst with an accuracy of 0.699

Recall - It measures how many observations out of all positive observations, have we classified as positive. Taking our customer churn example, it tells us how many churned customers we recalled from all the churned customers.

$$Recall = \frac{tp}{tp + fn} \quad \ldots (2)$$

While optimizing recall, you want to make sure you have identified ALL the customers who could churn.

For the SVM Classifier,

$$Recall = \frac{391}{391 + 345} = 0.798 = 79.8\%$$

For Decision Tree,

$$Recall = \frac{325}{325 + 324} = 0.455 = 45.5\%$$

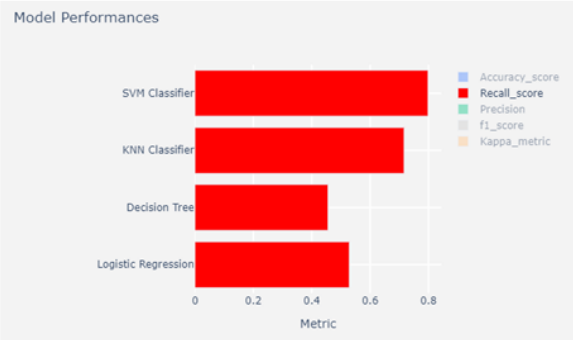

Figure 13. From the above figure, it is clear that SVM classifier has the highest recall score of 0.798 while the Decision Tree has the least recall score of 0.455

Model Performances

Logistic Regression

Figure 14. From the above classification report and ROC, the following information can be concluded:

Accuracy of 0.80 indicates that 80% of the customers were correctly classified.

Precision of 0.83 indicates that83% of the churned customers predicted by the model have actually churned.

Recall score of 0.91 indicates that the model was able to predict 91% of the actual churned customers as churned.F1 score of 0.87 out of maximum of 1 indicates that the model performs really well.

The Area under Curve (AUC) of the Receiver Operating Characteristic(ROC) is 0.71 out of a maximum of 1 which again indicates thatthe model has a good performance.

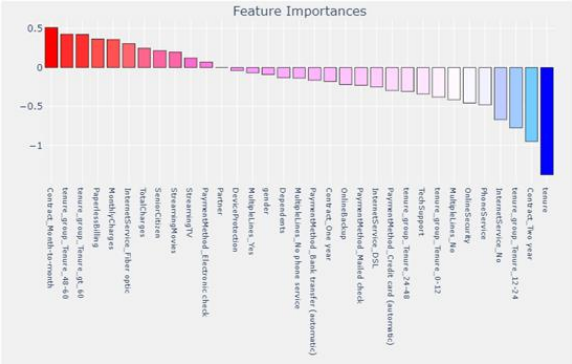

Figure 15. From the Logistic Regression algorithm, the above graph highlights the following points:

Attributes such as Contract_Two_year, Tenure_group_12-24 and InternetService_No contribute the most towards churn. This implies that customers having a two-year contract with the company, or who have stayed with the company for 12 to 24 months or have no internet service are more likely to leave the company. Attributes such as Contract_Month-to-month, Tenure_group_48-60, Tenure_group_gt_60 and PaperlessBilling contribute the least towards churn. This implies that customers having a monthly contract with the company, or who have stayed with the company for more than 48 months or have enrolled for a paperless billing service are more likely to stay with the company.

Attributes such as Partner, DeviceProtection and MultipleLines_Yes have negligible contribution in deciding customer churn. This implies that having a partner or not, or the device protection service or not, or multiple phone lines plays an insignificant role in estimating the likelihood of a custormer to leave the company.

Comparison of Models

A thorough comparison of algorithms based on the metrics mentioned above gives a comprehensive insight into the performance and efficiency of each of them. Theirperformances can be summarized as follows:

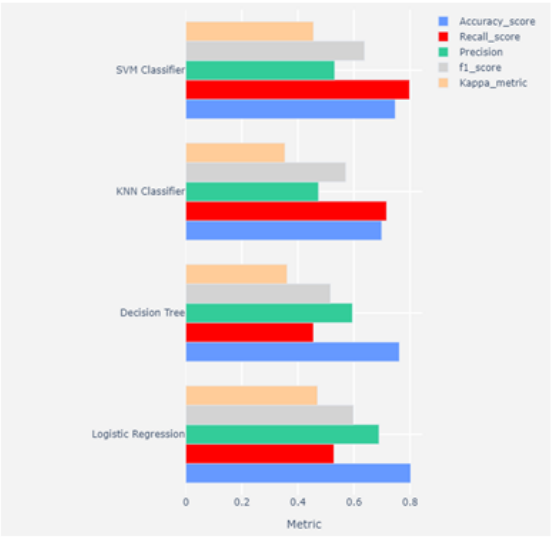

Figure 16. Graphical summarization of the performances of all the algorithms used

Table 1.Comparison of Results

Model	Accuracy_score	Recall_score	Precision	f1_score	Kappa_metric
Logistic Regression	0.8020	0.5286	0.6888	0.5982	0.4698
Decision Tree	0.7218	0.4551	0.5947	0.5156	0.3612
KNN Classifier	0.6991	0.7163	0.4737	0.5703	0.3532
SVM Classifier	0.7474	0.798	0.5312	0.6378	0.4557

From the above table, we observe that the results predicted by the Logistic Regression algorithm are the most efficient, evident from the high accuracy, precision, kappa metric and f1 score.

Conclusion

Churn prediction is one of the most effective strategies used in telecom sector to retain existing customers. It leads directly to improved cost allocation in customer relationship management activities, retaining revenue and profits in future. It also has several positive indirect impacts such as increasing customer's loyalty, lowering customer's sensitivity to competitors marketing activities, and helps to build positive image through satisfied customers.

The results predicted by the Logistic Regression algorithm were the most efficient with an accuracy of 80.2%. Therefore, companies that want to prevent customer churn should utilize this algorithm and remove features like long term contracts and instead replace them with monthly or short term contracts, thereby giving them more flexibility. Providing additional services such as device protection and multiple phone lines proves to be of little value to customer attrition. Lastly, focusing on enhancing the experience of loyal customers who have stayed with the company for long will prove worthwhile, ensuring their retention. The ability to identify customers that aren't happy with provided solutions allows businesses to learn about product or pricing plan weak points, operation issues, as well as customer preferences and expectations to proactively reduce reasons for churn.

Future Work

An important area for future research is to use a customer profiling methodology for developing a real-time monitoring system for churn prediction. Research dedicated to the development of an exhaustive customer loyalty value would have significant benefits to industry. It is anticipated that the profiling methodology could provide an insight into customer behaviour, spending patterns, cross-selling and up-selling opportunities. Seasonal trends could be apparent if the same data was studied over a period of several years.

A comparative analysis of prediction model building time with respect to different classifiers could be done in order to assist telecom analysts to pick a classifier which not only gives accurate results in terms of TP rate, AUC and lift curve but also scales well with high dimension and large volume of call records data. As concrete findings are related to the telecom dataset, other domains' datasets might be subject for further exploration and testing. Also, different and a greater number of performance metrics with respect to business context and interpretability might be explored in future.

BIBLIOGRAPHY

1. *Pavan Raj.Telecom Customer Churn Prediction, October 29th, 2018. Available: https://www.kaggle.com/pavanraj159/telecom-customer-churn-prediction/*
2. *Dataset resource link. Available: https://www.kaggle.com/blastchar/telco-customer-churn*
3. *Azeem, M., Usman, M. & Fong, A.C.M. A churn prediction model for prepaid customers in telecom using fuzzy classifiers. Telecommun Syst 66, 603–614 (2017).*
4. *Vijaya, J. & Elango, Sivasankar. An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. Cluster Computing. 22. 10.1007/s10586-017-1172-1 (2017).*
5. *Fridrich, Martin. Hyperparameter optimization of artificial neural network in customer churn prediction using genetic algorithm. 11. 9. 10.13164/trends.2017.28.9(2017).*
6. *Gordini, Niccolo & Veglio, Valerio. Customers Churn Prediction And Marketing Retention Strategies. An Application of Support Vector Machines Based On the Auc Parameter-Selection Technique In B2B E-Commerce Industry. Industrial Marketing Management. 62. 10.1016/j.indmarman.2016.08.003(2016).*
7. *Bahari, Tirani and M. Sudheep Elayidom. "An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour." (2015).*
8. *Kumar, Dudyala & Ravi, Vadlamani. Predicting credit card customer churn in banks using data mining. International Journal of Data Analysis Techniques and Strategies. 1. 4-28. 10.1504/IJDATS.2008.020020(2008).*
9. *Customer Churn Prediction for Subscription Businesses Using Machine Learning: Main Approaches and Models. Available: https://www.altexsoft.com/blog/business/customer-churn-prediction-for-subscription-businesses-using-machine-learning-main-approaches-and-models/*
10. *Hands-on: Predict Customer Churn. Available: https://towardsdatascience.com/hands-on-predict-customer-churn-5c2a42806266*