International Journal of Engineering & Technology, 7 (2.27) (2018) 69-76

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Predicting customer churn using targeted proactive retention

B. Mishachandar[1]*, Kakelli Anil Kumar[2]

[1]*M. Tech CSE, School of Computer Science and Engineering, VIT Vellore, India*
[2]*Associate Professor, School of Computer Science and Engineering, VIT Vellore, India*
**Corresponding author E-mail: mishachandar.b01@gmail.com*

Abstract

With the advent of innovative technologies and fierce competition, the choices for customers to choose from have increased tremendously in number. Especially in the case of a telecommunication industry, where deregulation is at its peak. Every year a new company springs up offering fitter options for its customers. This has turned the concentration of the business doers on churn prediction and business management models to sustain their places. Businesses approach churn in two ways, one is through targeted customer retention and through cause identification strategy. The literature of this paper provides a comprehensible understanding of the so far employed techniques in predicting customer churn. From that, it is quite evident that less attention has been given to the accuracy and the intuitiveness of churn models developed. Therefore, a novel approach of combining the models of Machine Learning and Big Data Analytics tools was proposed to deal churn prediction effectively. The purpose of this proposed work is to apply a novel retention technique called the targeted proactive retention to predict customer churning behavior in advance and help in their retention. This proposed work will help telecom companies to comprehend the risk associated with customer churn by predicting the possibility and the time of occurrence.

Keywords: *Big Data Analytics; Churn Prediction; Customer Churn; Machine Learning; Targeted Proactive Retention*

1. Introduction

1.1. Big data analytics

Big Data Analytics is a process by which we can uncover hidden patterns, unknown correlations, market trends and customer performances. It is also a process of cleaning, transforming and modelling the data in identifying new opportunities. Big Data analysis is mainly used for product promotion, faster and better decision making and cost reduction [1]. It also has application in healthcare, finance, crime detection, public sector and gaming etc.

1.2. Churn

The churn in the customer churn prediction refers to the customer act of culminating his or her relationship with the service provider due to disgruntlement over its services [2]. Acts such as unsubscribing a service, uninstalling an application and deactivating are all examples of customer churn. Over the last few decades, the telecom industry has witnessed enormous developmental changes in terms of increase in the level of competition and the competitors, openings to new services and the burgeoning technology industry. A churn of a customer severally hits the company's revenue and its marketing expenses [3].
A Churn is an evident result of customer's discontent over the company's services. Identifying the root cause for this dissatisfaction demands the need for a set of parameters. A customer wholly withdrawing form a service doesn't happen in a day, rather a long-term uncontacted and displeased behavior from the service provider results in such impetuous gesture from the customer's end. To overcome such behavior, the service provider maintains a detailed status report of the customer's mode of operation to apprehend their status and to predict their longevity in continuing the services.

1.3. Churn prediction

To predict that a customer is likely to churn, there is a great additional potential revenue source for all the online business. In addition, to loss of revenue, the customer may not have already been covered by customers spending to date. Furthermore, it is always more difficult and expensive to acquire a new customer than it is to retain a current paying customer [4].
Churn prediction is a humongous business. Its main aim is to predict the customer's inactivity, precisely the ones which are likely to end a subscription to a service [5]. It is mainly deployed in the telecommunications industry, where the interaction between the customers and the service renders is comparatively high. Apart from it, it also finds use in various other domains like banks, ISPs, insurance organizations and so on.

1.4. Causes of customer churn

The root cause of customer churn is dissatisfaction and displeased behavior from the service providers. Acts such as unsubscribing a service, uninstalling an application and deactivating are all examples of customer churn. Leading cause of churn is poor customer service. Poor customer services and dissatisfaction can, therefore, lead to many such customers churning in the future.
If the service provider is unaware of a customer who is about to churn, no action can be taken for that customer. Business must consider risk, the level, and cost of intervention and plausible customer segmentation [6]. In telecommunication industry, subscribers frequently switch over from one industry to another which is a prime concern.

1.5. Types of churn

1) Proactive churn

Customers who proactively cancel their subscription themselves, which is a bad churn and must be minimalized.

2) Reactive churn

When customers forget to update their credit card information, the account gets canceled. By a solution called Dunning Management, customers can be automatically informed when their charge was declined.

3) Happy churn

When a customer cancels your product after using it for their campaign, for a specific time frame without the need for a constant usage.

4) Fake churn

Many SaaS companies have a 30 to 90 money back guarantee available which is a good measure to understand how much it is costing you.

1.6. Importance of churn prediction

Customer churn prediction has become the need of the hour in telecom sector due to the hasty hike in customer count and telecom companies. The prime focus of any telecom company now a day is to hold back their long-term customers instead of acquiring new ones, to fetch high profits [7]. The reason for spotlighting on prevailing customers is mainly to ameliorate the standard of sales and scale down the linked marketing cost as to their new customers. Factors such as these have created a necessity for churn prediction activity to grip as a fundamental part of the telecom sector.
Handset or device choice is a well-known driver of churn in the mobile phone business. As a result, a popular policy is to subsidize the price of the handset of new subscribers and charge full price to existing customers to upgrade. This policy has led to customers hopping from one provider to another to get a new discount [8]. This, in turn, has prompted providers to redefine their strategies.

1.7. Limitations of churn prediction

High volatility in handset offering is a factor that quickly invalidates the models of churn on current handset models [9]. The net result for modelling is that we cannot devise a sound policy simply by eliminating no reasons for churn. In fact, a continuous modelling strategy including classic models that quantify categorical variable is mandatory.

2. Motivation and significance

The literature review of this paper speaks on the significance of customer retention over acquiring fresh customers. Retaining existing customers is significantly more reliable and cost-effective. In the current competitive business scenario with a plunge in the business doers, the effect of losing even a single customer is perilous. Loss of a single customer in a flourishing business firm can be viewed from distinct aspects. Firstly, every customer associated with a business is considered an asset. Losing such a customer is equivalent to irreparably damaging a vital functioning machine in the company [10]. Likewise, in the same hypothetical assumption, loss of a customer is equivalent to conceding the business to our peer competitors. Lastly, acquiring a new customer is no easy business. It demands a huge effort from the company's end. And even on doing so, there is no guarantee that the customer will stay long with the business. Winning even the loyalty of a smaller number of customers is uncertain [11]. Hence, the prevention approach has been quite worthwhile. Customer Retention is of immense importance in enterprises like the matured telecom and finance company. Achieving it demands a pertinent churn prediction model, which is another similar terminology associated with customer retention. Customer churn prediction can be interpreted as predicting the customer's behavioral tendency to switch over to a new company.
The today's fast progressing business habitat especially in telecom industries, the level of competition is alarming. The want for better services and varied options has also made the competition more intense. That is the reason where customers' loyalty becomes a question. Lack of such loyalty creates an unstable tendency in customers to easily shift from one provider to another. After all, it's a matter of choice for the customers to decide on what service they wish to adopt. Now it is the task of the service providers to work on the customer felt limitation to retain them. Therefore, it is obligatory to segregate the customers as churn and non-churn. Non -churn refers to the set of customers who are reluctant to shift from the service of one provider to another in contrast to churn customer. This has created the practice of churn prediction a must in the telecommunications industry. The telecom companies are the ones that get drastically affected by churners. The loss in terms of customers, money and time is humungous [12]. Due to the changes and measures taken in the recent times, the count has receded and is marked as a genuine issue.

3. Limitations of existing churn prediction systems

Churn is a metric that affects marketing. Machine learning and survival analysis are the two approaches for churn modelling. Machine learning methods are widely used due to their high performance and ability. On the other hand, survival analysis can provide value by answering a distinct set of questions. In addition to these two approaches, Ensemble models can provide superior accuracy but time-consuming [13]. A churn model is only as good as the future going into it. In addition to domain knowledge, skill and creativity are needed to construct a robust future that is predictive of a churn event. Careful exploratory analysis and sometimes auxiliary model building often must occur before you embark on building an overall churn model.

3.1. Customer retention

Retaining customers in the telecom industry is becoming challenging each day [14-15]. Therefore, a continuous modelling is mandatory. Churn detection on big data sets on the customers is an effective approach.
Customer churn is indirectly proportional to the growth of a company. As such, the greater your customer churn rate, the lesser chances of growing business. Even if you have the fittest of marketing campaigns and strategies in your industry, your bottom line suffers if you are losing customers at a large rate, as the cost of acquiring new customers is relatively high. Customer acquisition cost always exceeds the customer retention costs [16]. Companies spend seven-times more acquisition than customer retention.
Therefore, customer churn is costly for business. To assist the workflow of this proposed work, a novel idea of deploying Azure to perform the functions of big data analytics is used with modelling techniques like classification and decision tree. Finally, the experimental results reveal that Azure one of the effective option for predicting churn.

4. Literature survey

The prior focus of any customer concerned company is to retain its customers [21]. To aid the Customer Relationship Management (CRM), a discrete set of tools is developed. This invariably guides in the prediction and classification models. In an era of fast businesses, creating abiding bonds with customers and keenly noting their behavior around the clock are key boost points to flourish any business [18]. Recent researchers have turned their attention towards various machine learning (ML) techniques like Random Forest, RotBoost, Support Vector Machine (SVM) to deal with churn prediction [19,20]. A balanced combination of Big Data Analytics and discrete supervised machine learning algorithms contributes to a well-structured framework to overrule the problems concerning the prediction of churn.

4.1. Comparative study of existing churn prediction systems

Table 1: Comparative Study of the Existing Churn Prediction Systems

Author	Citation	Title	Year	Aim	Technique	Records/Training Testing	Dataset	Outcome
Bingquan Huang, Mohand Tahar Kichadi, Brian Buckley	[1]	Customer Churn Prediction in Telecommunication Industry	2011	Landline customer churn prediction	Logistic, Regression, Linear classification, Naïve Bayes, Decision tree, SVM, Multilayer Perceptron Neural Network, Evolutionary Data Mining Algorithm	13,562 churners & 400,000 non-churners	Real world datasets collected from the telecoms of Ireland	New features compared to the six modelling techniques is more effective.
Michael C. Mozer, Richard Wolniewicz, David B. Grimes, Eric Johnson, Howard Kaushansk	[2]	Predicting Subscriber Dissatisfaction and Improving Retention in the wireless telecommunications industry	2000	To predict customer churn and based on the incentives of the subscribers is decided.	Logistic Regression, Decision Tree, Neural Networks and Boosting Algorithm	46,744 business subscribers	Wireless carrier	A monthly churn rate of 3.1%. was achieved.
Adnan Amina, Sajid Anwara, Awais Adnana, Muhammad Nawaza, Khalid Alawfib, Amir Hussainc, Kaizhu Huang	[3]	Customer Churn Prediction in Telecommunication sector using Rough Set Approach	2016	To extract important set of decision rules based on churners and non-churners.	Rule based decision making technique-Rough set approach	50% Instances for training and 50% testing	Bench mark telecom dataset	Optimal solution obtained for CPP
Wael Etaiwi et al	[4]	Evaluation of Classification Algorithms for Banking Customer's Behaviour Under Apache Spark Data Processing System	2017	Evaluating classification algorithm to predict the natural customer's behavior in banks.	Apache spark Data processing system	13 million records for training and 1 million for training	Customer's personal and behavioral information in Santander Bank in Spain	Naïve Bayes overcomes SVM in terms of precision, recall and F-measure.
Ammar A Q, Ahmed, Maheshwari	[5]	Churn Prediction on Huge Data using Hybrid Firefly Algorithm	2017	To predict Customer Churn prediction in telecom company	A novel hybrid firefly classification Algorithm	80: 20 ratios	Orange dataset	Firefly algorithm works best on churn data and hybrid provides fast and efficient result.
Dries Harnie et al	[6]	Scaling Machine Learning for Target Prediction in Drug Discovery using Apache Spark	2015	Target prediction for drug discovery	Scaling machine learning using Apache spark	330,000 compounds and 3,000 targets	Real time disease dataset	In Spark pipeline vs original, pipeline is proved efficient.
Junxiang Lu	[7]	Predicting Customer Churn in the Telecommunications	2012	The aim of this paper is to apply Survival Analysis(SA) techniques for churn prediction.	Conventional statistical methods like logistics regression and decision tree	A dataset with 41,374 active high-value customer base data	Telecom database	To find Customer survival function was estimated and to demonstrate how survival techniques are used.

A keramati, R Jafari Ma-randhi,I Almadian	[8]	Improved churn prediction using Data mining tech-niques	2014	To predict churn using Data mining algorithms	Data mining classification techniques like SVM, ANN and K-nearest neighbor-hood	2:1 ratio	Dataset of Iranian mobile company	95% accuracy achievable through this method.
Yaya Xie , Xiu, Li Ngai, Weiyun Ying	[9]	Customer Churn Prediction using Improved Balanced Random Forests	2015	To propose improved a version of balanced random for-ests (IBRF) and exhibit its applica-tion.	IBRF Algo-rithm	20,000 cus-tomer's data with 27 varia-bles.	Real bank's data warehouse Dataset	IBRF was found to improve pre-diction accuracy significantly compared with other algorithms, like artificial neural networks, decision trees, and class-weighted core support vector machines (CWC-SVM)

5. The proposed work

5.1. Methodology

We designed a predictor by choosing a model class (Naïve Bayes classification, decision tree) in Microsoft Azure Workbench environment. The predictor in a local environment was tested with varied arguments ranging from 0.01 to 10 for every argument, a new job was executed, and the result produced showed the difference in accuracy rates. In our proposed work, we focused only on voluntary churn and not on involuntary churn. In voluntary churn, the customer decides to culminate his services with the service provider. Whereas in an involuntary churn, the service provider culminates its association with the customer due to bill arrears.

5.2. Targeted proactive retention

Though most of the mobile operators use proactive retention start for preventing subscribers churn, they were unable to succeed in their attempts to do so. Portability of mobile numbers can be done by subscribers within a short period of time. However, churn issue will continue to rise amongst the operators around the world.
Mobile operators offer incentives like a yearlong contract, cost-free call for few minutes for a certain number of months. Not only to retain the subscribers but also stem churn.
The above incentives work to some extent for the post-paid subscribers, but it is not effective for the prepaid subscribers who dominate 90% of the market [22]. Before subscriber ports out, mobile operators should address churn by proactive retention.
Steps to set up proactive retention strategies

1) Define churn model for both prepaid and post-paid subscribers in building a proactive retention practice.
2) Build a predictive churn model is a tool that helps Telecom Company to identify which of its subscribers will churn to prevent it from happening.
3) Design offers for the various groups predicted to churn based on the results of the predictive churn model. The offer could be a certain number of free minutes each month for a prescribed number of months and discount on a fixed post-paid plan to loyal subscribers.
4) The final step is the piloting phase for examining the examiners of the materials as well as agents for the retention efforts for the churners. The challenge of retaining existing subscribers is drastically becoming more and more troublesome.

5.3. Four step implementations

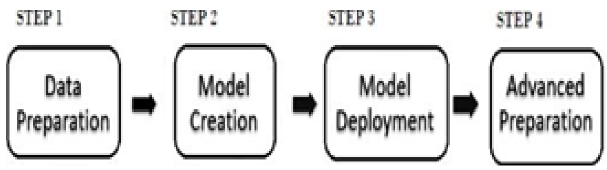


Fig. 1: Four-Step Implementations.

1) Data preparation phase

In the data preparation phase of the proposed work, once the prerequisites are met, the implementation begins by creating a new project in the Azure Machine Learning Workbench environment. Followed by which the data from the telecommunications industry is created as a separate data preparation package. Azure Workbench has inbuilt packages like Python, spark, scikit-learn, and matplotlib to aid easier scripts execution. In this case, the Python/PySpark code is generated to invoke the data preparation package.

2) Model creation phase

The model creation phase is the important stage in the proposed work. It is when the prediction model gets its form. Once the initial phase of data preparation is over, the Azure environment and the dataset is now ready for Machine Learning to be applied. The code for the model created in done Python. Azure facilitates various environments for script execution namely, Local environment, Local Docker environment and Local Azure CLI window. Our project is executed only in the local environment. Scripts execution in Local Docker environment requires Docker engine to be installed and started locally on the system. Similarly, to run scripts against Local Azure CLI window, a remote Azure VM or an Azure HDInsight Spark cluster should be created.

3) Model deployment phase

Post model creation, the appropriate model file is located. In this case "pickle "is the model file. The pickle module of Python is an important algorithm for serializing and de-serializing a Python object structure. After choosing the appropriate model file a scoring script and a schema file are generated. Before starting with the web service creation, the environment is prepared, and a real-time web service is created and executed. The result of the web service is examined in the Azure blob storage as blob data.

4) Advanced data preparation phase

In the advanced data preparation phase, firstly the data is prepared with the ML data preparation tool. The prepared data is then imported, transformed to create a test dataset. Once the data is prepared, a data package for preparation is generated and executed using Python. For more input files a training set of data is generated by reusing the data preparation package. Finally, the scripts are executed in the local Azure CLI window and Cloud Azure HDInsight environment.

6. Architectural diagram

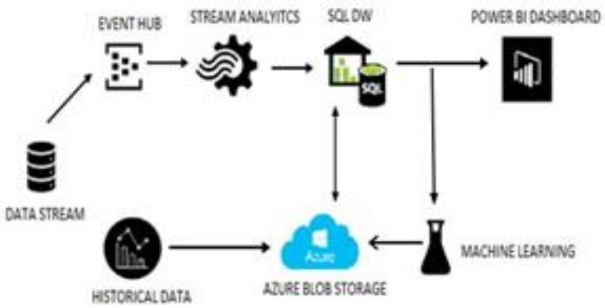


Fig. 2: Architectural Diagram.

7. Algorithm

7.1. Naïve Bayes classification algorithm

Naive Bayes algorithm is the most preferred and a popular classification algorithm [23-24]. It is derived from the Bayesian theorem of classification. It is best suited for problems involving high dimensional inputs.
Parameter estimation for naive Bayes models uses the method of maximum likelihood. In spite over-simplified assumptions, it often performs better in many complex real-world situations. Advantage: Requires a small amount of training data to estimate the parameters.

7.2. Bayes theorem

$$P(c|x) = P(x|c)\ P(c) \quad (1)$$

P(x)

Definition 7.2.1: *P (c|x) denotes posterior probability of class (c, target) given predictor (x, attributes).*

$$P(x) \quad (2)$$

Definition 7.2.2: *P(c) denotes prior probability of class*

$$\times (x2|\ c) \times \ldots.P(xn|c) \times \quad (3)$$

Definition 7.2.3: *P (x|c) denotes likelihood*

$$P(c|X) = P(xi|c) \quad (4)$$

Definition 7.2.4: *P(x) I - prior probability of predictor*

7.3. Applications of naive bayes classification

1) Naive Bayes algorithm finds application in classifying textual contents.
2) It is practically used for filtering spam in emails.
3) A Hybrid form of Recommender System can be created using the Naive Bayes Classifier and Collaborative Filtering.
4) It is a preferred choice of algorithm for designing applications online.

7.4. Decision tree algorithm

Decision tree represents the models of classification and regression in a tree structure [25]. The dataset is fragmented into smaller subsets at each iteration which then leads to a decision tree finally. The result is a decision tree with decision nodes and leaf nodes. The decision node in the decision tree has more than two sub-branches and the leaf node denotes a classification or decision. The first top node in the tree is called root node and it represents the best predictor. Both categorical and numerical data can be dealt using DT.

8. Experimental result investigation/analysis

8.1. Introduction to experimentation

The entire experiment is performed on the Azure Workbench (Preview) version with Dockers installed explicitly. Azure workbench is compatible only with Windows 10 and 10 Pro. A pre-requisite of a Microsoft account and an Azure account is a must. The four stages of implementation have already been discussed in the methodologies section of this paper. The choice for the dataset is left to the user as Azure is compatible with any size of dataset ranging from small to medium to large to very large. The necessary packages to run a Python code are all installed in the workbench by default. This entire proposed work can be sufficiently performed in the one-month free trial version of workbench itself. But to enable few advanced features, a paid version will help. The hardware and software specifications of the system used for experimentation is mentioned in the section below.

9. System requirements specifications

The experiments were conducted on a Windows 10, 64 Bit system. The experiments are compatible on both Mac OS and Windows 10 or Windows 10 Pro. For the software requirements, a Microsoft account (mandatory) with Microsoft Azure free or paid account is needed to deploy the services on cloud. A Community version of the Docker is required for performing operating-system-level virtualization also known as containerization. A Docker Toolbox can be installed to setup a quick launch of the Docker. The toolbox consists of the Docker Quick start Terminal, Oracle VM VirtualBox and Kinematic.

9.1. Microsoft azure

Microsoft Azure Machine Learning service is an integrated, end-to-end analytical solution to aid data scientists in data preparation, experimental development, and model deployment at the public cloud scale.
Machine Learning, a data science technique that permits computers to utilize the prevailing data to perform future forecasting of behaviour, outcomes, and trends. Computers acquire the ability to learn even without being programmed explicitly using Machine Learning. Tasks like prediction and forecasting from ML paves way for the invention of fitter applications and devices. Product recommendation based on the items in your cart during online purchases is made possible using ML. when a credit card is being swiped during a transaction, the current transaction is analogized with the previous set of transactions to detect the chances of fraudulence. This is again possible only with Machine Learning. Similarly, in the case of a robot enabled vacuum cleaner, ML helps in directing the robot to get its work done. Hence, ML is now trending as the game changer in most technological gadgets making the field of Data science, the most demanding.

9.2. Machine learning in Microsoft azure cloud

Azure ML is a cloud-enabled predictive analytic service that easies model creation and deployment effortless in model predictive analytics.

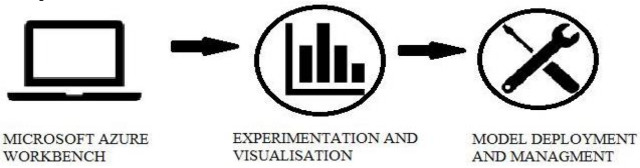


Fig. 3: Azure Machine Learning.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	age	annualinc	calldropra	callfailure	callingnur	customeri	customers	education	gender	homeown	maritalsta	monthlyb	noadditio	numberof	numberof	numdaysc	occupatio	penaltyto	state	totalmins	unpai
2	12	168147	0.06	0	4.25E+09	1	Yes	Bachelor (	Male	Yes	Single	71	\N	0	7	96	Technolog	371	WA	15	
3	12	168147	0.06	0	4.25E+09	1	Yes	Bachelor (	Male	Yes	Single	71	\N	0	7	96	Technolog	371	WA	15	
4	42	29047	0.05	0.01	4.25E+09	2	Yes	Bachelor (	Female	Yes	Single	8	\N	1	4	14	Technolog	43	WI	212	
5	42	29047	0.05	0.01	4.25E+09	2	Yes	Bachelor (	Female	Yes	Single	8	\N	1	4	14	Technolog	43	WI	212	
6	58	27076	0.07	0.02	4.25E+09	3	Yes	Master or	Female	Yes	Single	16	\N	0	2	55	Technolog	403	KS	216	
7	58	27076	0.07	0.02	4.25E+09	3	Yes	Master or	Female	Yes	Single	16	\N	0	2	55	Technolog	403	KS	216	
8	20	137977	0.05	0.03	4.25E+09	4	Yes	PhD or eq	Male	No	Single	74	\N	1	7	73	Technolog	76	KY	412	
9	20	137977	0.05	0.03	4.25E+09	4	Yes	PhD or eq	Male	No	Single	74	\N	1	7	73	Technolog	76	KY	412	
10	36	136006	0.07	0	4.25E+09	5	Yes	High Scho	Male	Yes	Married	81	\N	0	5	14	Technolog	436	ND	416	
11	36	136006	0.07	0	4.25E+09	5	Yes	High Scho	Male	Yes	Married	81	\N	0	5	14	Technolog	436	ND	416	
12	67	246906	0.05	0.01	4.25E+09	6	Yes	High Scho	Female	Yes	Married	19	\N	1	2	32	Technolog	108	OK	113	
13	67	246906	0.05	0.01	4.25E+09	6	Yes	High Scho	Female	Yes	Married	19	\N	1	2	32	Technolog	108	OK	113	
14	14	244935	0.07	0.02	4.25E+09	7	Yes	High Scho	Male	Yes	Single	27	\N	3	0	73	Technolog	468	AZ	116	
15	14	244935	0.07	0.02	4.25E+09	7	Yes	High Scho	Male	Yes	Single	27	\N	3	0	73	Technolog	468	AZ	116	

Fig. 4: Dataset.

Azure Machine Learning renders both tools and inbuilt packages for model predictive analytics and a set of fully managed services that help in converting predictive models to a functioning web service.

Predictive analytics makes use of algorithms that analyse historical or current prevailing data to help in detecting patterns or trends to forecast events in the future. Tools are designed and provided to build a complete set of machine learning solutions in the cloud environment.

Azure ML renders a complete modelling set consisting of a huge algorithm library, a work studio for model creation and web services deployment to develop a full-fledged predictive analytics solution in the public cloud.

9.3. Azure machine learning workbench

Azure Machine Learning Workbench is a client application that runs either as a desktop application or command-line tool for both Windows and MacOS. Any Machine learning solution can be managed through the vast data science life cycle of Azure.

Data science life cycle consists of three pre-dominant phases namely:

1) Data ingestion and preparation
2) Model development and experiment management
3) Model deployment in multi-targeted environments

Core functionalities of Azure Machine Learning Workbench:

1) Workbench facilitates Data preparation which makes learning data transformation logic with appropriate examples easier.
2) Provides various built-in services like client UX, Jupyter Notebook etc.
3) Overall run history helps in monitoring the experiment.
4) Inbuilt Python and Spark package aid data source abstraction.
5) Autosaving option

9.4. Characteristics of input data

The root cause of churn is customer dissatisfaction that eventually gets converted to disloyalty towards the service provider. The dissatisfaction can be due to various evident reasons like a compromise in the quality of service, high costs or even time delay [26]. The reason for customer discontent is not always the same but vary over time and interests. The input data in our proposed work is categorized for better understanding as, Demographic Data, Usage level data, Quality of service data and other features/marketing data.

1) Demographic data: Data relating to the population and geography of a region.
2) Usage level data: Details of customer call history like call duration, caller location, calls count, call limit, etc.
3) Quality of Service (QoS) data: As its name stands for, this data deals mostly with the call experience like quality, network coverage, connection interruption, voice clarity, etc.
4) Other feature / Marketing data: Data relating to promotion fall into this category like SMS, emails, advertisements, new competition, call tariffs, etc.

9.5. Dataset

The dataset used in our proposed work is a very large sample customer dataset of 20,469 records and 26 attributes containing basic customer information like Annual Income, Call Drop Rate, Call Fault Rate, Calling Number, Customer ID, Customer Usage, Education, Gender, Home Owner, Marital Status, Monthly Billed Amount, Number Of Additional Lines, Number Of Complaints, Expiration Period, Occupation, Penalty To Switch, State, Total Minutes Used In Last Month, Unpaid Balance, User Internet Service, User Voice Service, Percentage Call Outside Network, Total Call Duration, Average Call Duration, Churn, Year and Month. The important attribute of the dataset is the churn. The dataset consists of data that was recorded for a period of 2 months of the year 2015.

9.6. Training and testing in azure environment

The training and testing procedure of data in Azure is performed by Analysis Services, which haphazardly samples the input data and maintains a similarity between the testing and training sets. Managing similarity between the sets aids in minimizing the effects of data discrepancies and paves way for better interpretation of the model.

As mentioned earlier, the churn data is the key attribute. Azure makes training and testing easier using the Data Mining Wizard which by default, divides the dataset into training and testing data. The training of the model is performed using churn records and the testing function by the non-churn records. A default of 70:30 ratio is the most optimum in data mining, but the ratio can also be altered matching our requirements using Analysis Services. Post model creation phase, the training dataset is used to process the model and the testing is done by making a prediction against the test data. The test data contains all the possible prediction values for the attribute. This feature of testing procedure aids model guessing easy.

The proposed work is designed in such a way that it is compatible with any type of data ranging from big to small and trained to untrained. This is explained with appropriate results in the results and discussions section of this paper.

9.7. Performance metrics

Many measures can be used to judge the performance of prediction algorithms, entropy, purity, true positive and negative rate, accuracy, precision, F-Measure and computation time [27].

The performance metrics used for valuation are precision, recall, and F-Measure. The measure can be calculated using the value of both precision and recall. Accuracy, Sensitivity, Specificity, training period and prediction time are used as evaluation metrics to measure the performance of the algorithms.

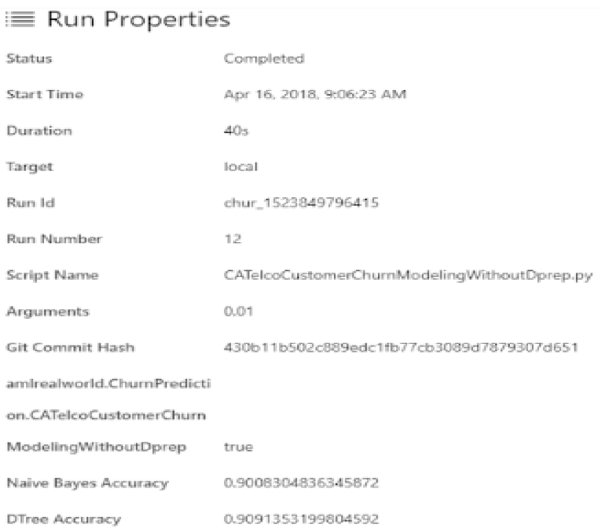
With customer's success, we are proactively helping customers towards goals whereas, with customer support, we are relatively fixing things and answering questions.

It is easier to drive revenue through upselling and counselling existing customers versus going out and finding new customers. 90% of revenue is derived from customer success. The other major benefit of investing in customer success is getting new business through virality and word-of-mouth. Therefore, we need to have customer success metrics in place that help evaluate our progress. Expansion revenue from the existing customer is a measure of how good we are getting customers to grow with our product by periodically monitoring customer's contentment.

While the customers at the brim of churning can be spotted and efforts to prevent it can be taken by enhancing the services [29].

Customers stick around our company longer; spend more but less expensive to save and tell other people about our product. Customer's satisfaction is paramount for the growth of product which predicts their loyalty towards the company.

10. Result screenshots



Run Properties

Status	Completed
Start Time	Apr 16, 2018, 9:06:23 AM
Duration	40s
Target	local
Run Id	chur_1523849796415
Run Number	12
Script Name	CATelcoCustomerChurnModelingWithoutDprep.py
Arguments	0.01
Git Commit Hash	430b11b502c889edc1fb77cb3089d7879307d651
amlrealworld.ChurnPrediction.CATelcoCustomerChurnModelingWithoutDprep	true
Naive Bayes Accuracy	0.9008304836345872
DTree Accuracy	0.9091353199804592

Fig. 5: Naïve Bayes and D Tree Accuracy.

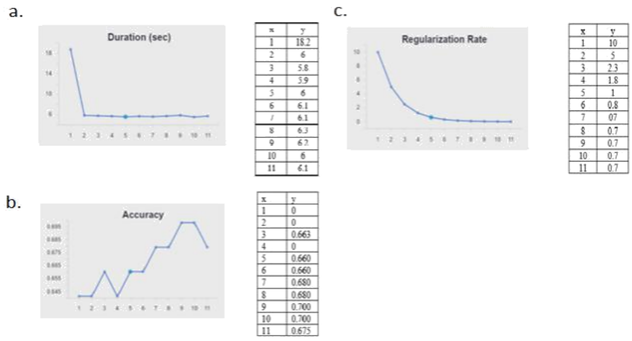
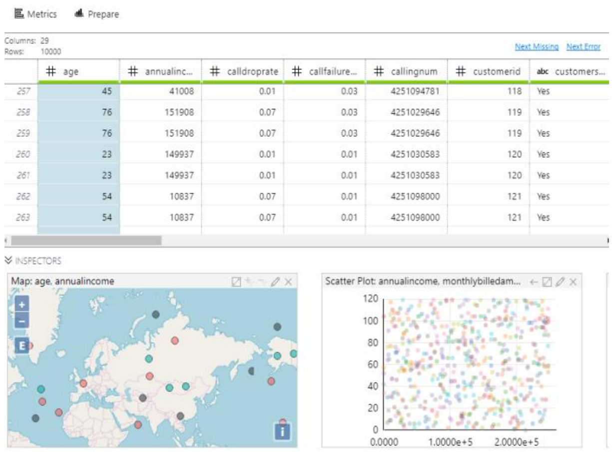


Fig. 6: A) Duration Graph Indicating the Time Taken for the Complete Job Execution. B) Accuracy Graph Indicating the Accuracy Achieved During Multiple Jobs Execution. C) Regulation Graph Indicting the Changes in Regulation Rate at Various Time.



	# age	# annualinc...	# calldroprate	# callfailure...	# callingnum	# customerid	abc customers...
257	45	41008	0.01	0.03	4251094781	118	Yes
258	76	151908	0.07	0.03	4251029646	119	Yes
259	76	151908	0.07	0.03	4251029646	119	Yes
260	23	149937	0.01	0.01	4251030583	120	Yes
261	23	149937	0.01	0.01	4251030583	120	Yes
262	54	10837	0.07	0.01	4251098000	121	Yes
263	54	10837	0.07	0.01	4251098000	121	Yes

Fig. 7: Metrics form of the dataset

11. Result analysis

11.1. Churn prediction

For every individual predictor, an estimate of the probability of churners is obtained by combining the three output graphs namely, Accuracy, Regulation rate and Duration [28]. The steeper the graph, the better the accuracy rate. Different sets of graphs are obtained for the different regulation rate that is inputted. But finding whether a customer is a churn, or no churn will be expected result of this proposed work. This is obtained by finding a threshold value from the list of continuous probabilistic outputs. The output to present the churn and no churn are always expressed in the form of a lift curve in most telecom industries for easier visual interpretation. The lift curve is presented in the form of a multi-class ROC depicting ROC for every area in the graph. The graph is plotted against churners (above threshold) vs non-churners (below threshold) to determine the churners probability. The lift curve plots one ROC quantity against the other.

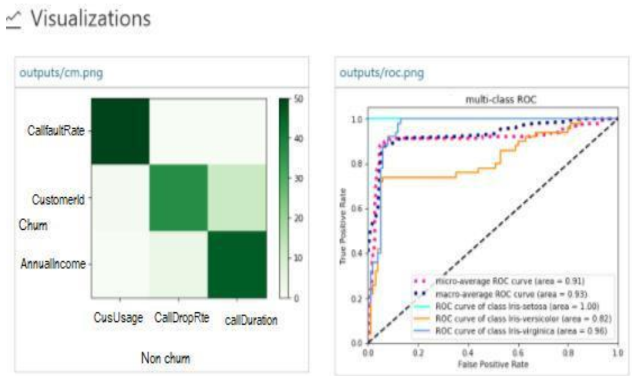


Fig. 8: Multi-Class ROC Graph.

12. Conclusion

Churn prediction with retention strategies is becoming a must-have in any industry that is facing fierce competition from its peer business doers. To deal with proactive churners, an optimum churn prediction model alone is not sufficient. A prediction model with an efficient retention strategy is the right combinational model for predicting proactive churners in the present day competitive environment. Therefore, to design and build such a model with a desired level accuracy and precision has become a research constraint for researchers in the recent times. In this proposed work, we considered two different novel Machine Learning classification-based algorithms to perform churn prediction. In specific, Naïve Bayes Classification algorithm and Decision tree algorithm were considered. Machine Learning Modelling in combination with Big Data Analytics tool for implementation can create a desirable environment for predicting churn. In no sooner businesses will prefer Churn prediction with retention strategy as a prevention approach than as a prediction model.

12.1. The ongoing research

The research part of the proposed work will proceed in different directions until the most optimum churn model is designed. Achieving the desired accuracy is the expected result of this proposed work. As part of the future work, the proposed work will be extended in different directions with a newer set of parameters. First, the churn model will be tested with distinct types of datasets (large to small) to check if changes in dataset impact the output. Second, the model can be tested with other ML algorithms to increase the efficiency of the model. Finally, to further improve the output accuracy different customer retention techniques can be tried. Therefore, churn prediction is a vast concept and is still unexplored in distinct aspects. The need for a better churn prediction model is always the want of any business, especially in the telecommunications industry [30]. So, the scope for future research is limitless in this field.

Acknowledgment

We thank Capt. S. Manohar for his assistance and comments that greatly improved the manuscript and VIT, Vellore for providing us the required support to try and execute the various techniques and methodologies for the successful completion of this project.

References

[1] Keramati A., Jafari-Marandi R., Aliannejadi M., Ahmadian I., Mozaffari M., Abbasi U. (2014), "Improved churn prediction in telecommunication industry using data mining techniques",Applied Soft Computing Journal, 24 , pp. 994-1012.
[2] Michael C. Mozer, Richard Wolniewicz, David B. Grimes, Eric Johnson, and Howard Kaushansky," Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry". IEEE Transactions on Neural Networks, 11:690–696, 2000.
[3] Adnan Amin, Sajid Anwar, Awais Adnan, Muhammad Nawaz, Khalid Alawfi, Amir Hussain and Kaizhu Huang, "Customer Churn Prediction in Telecommunication Sector using Rough Set Approach",Neurocomputing,http://dx.doi.org/10.1016/j.neucom.2016.12.009
[4] Etaiwi, W., Biltawi, M., Naymat, G,"Evaluation of classification algorithms for banking customer's behavior under apache spark data processing system" Procedia Comput. Sci. 113, 559–564 (2017)
[5] A. A. Ahmed, and D. Maheswari, "Churn prediction on huge telecom data using hybrid firefly-based classification," Egyptian Informatics Journal, 2017.
[6] Harnie, D., Vapirev, A.E., Wegner, J.K., Gedich, A., Steijaert, M., Wuyts, R., & De Meuter, W. (2015)," Scaling machine learning for target prediction in drug discovery using apache spark" In Proceedings of the 15th IEEE/ACM International Symposium on Cluster Cloud and Grid Computing.
[7] Junxiang Lu, "Predicting Customer Churn in the Telecommunications".
[8] Keramati A, Jafari-Marandi R, Aliannejadi M, Ahmadian I, Mozzafari M, Abbasi U (2014), "Improved churn prediction in telecommunication industry using data mining techniques", Appl Soft Comput 24:994–1012.
[9] Xie Y, Li X, Ngai EWT, Ying W (2009), "Customer churn prediction using improved balanced random forests" Expert Syst Appl 36(3):5445–5449
[10] Au, W., Chan, C., & Yao, X (2003), "A novel evolutionary data mining algorithm with applications to churn prediction*", IEEE Transactions on Evolutionary Computation*, 7, 532–545. https://doi.org/10.1109/TEVC.2003.819264.
[11] Boser, B., Guyon, I., & Vapnik V (1992), "A training algorithm for optimal margin classifiers", *In Proceedings the fifth annual ACM workshop on computational learning theory*, (pp. 144–152).
[12] Pittsburgh, PA: ACM Press. Bradley, A. P (1997), "The use of the area under the roc curve in the evaluation of machine learning algorithms", *Pattern Recognition*, 30, 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2.
[13] Burges, C. J. C (1998), "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, 2(2), 121–167. https://doi.org/10.1023/A:1009715923555.
[14] Coussement, K., & den Poe, D. V (2008), "Churn prediction in subscription services: An application of support vector machines while comparing two parameter selection techniques", *Expert Systems with Applications*, 34, 313–327. https://doi.org/10.1016/j.eswa.2006.09.038.
[15] Domingos, P., & Pazzani, M. J (1997), "On the optimality of the simple bayesian classifier under zero-one loss", *Machine Learning*, 29(2–3), 103–130. https://doi.org/10.1023/A:1007413511361.
[16] Hadden, J., Tiwari, A., Roy, R., & Ruta, D (2006), "Churn prediction: Does technology matter", *International Journal of Intelligent Technology*, 1(2).
[17] Henley(2009)., Hosmer, D., & Lemeshow, S (1989), "Applied logistic regression" , *New York: Wiley.*
[18] Huang, B., Kechadi, M.-T., & Buckley, B (2010), "A new feature set with new window techniques for customer churn prediction in landline telecommunication", *Expert Systems with Applications*, 37(5), 3657–3665. https://doi.org/10.1016/j.eswa.2009.10.025.
[19] Hung, S.-Y., Yen, D. C., & Wang, H.-Y (2006), "Applying data mining to telecom churn management", *Expert Systems with Applications*, 31, 515–524. https://doi.org/10.1016/j.eswa.2005.09.080.
[20] Japkowicz, N (2000), "Learning from imbalanced data sets: A comparison of various strategies", (pp. 10–15).
[21] AAAI Press. Japkowicz, N (2006), "Why question machine learning evaluation methods", *In AAAI Workshop*.
[22] Boston. John, H., Ashutosh, T., Rajkumar, R., Dymitr, R. (2007) "Computer assisted customer churn management: State-of-the-art and future trends"
[23] Jolliffe, I. T (1986), "Principal component analysis",*New York: Springer https://doi.org/10.1007/978-1-4757-1904-8*.
[24] L. Xi, Y. Wenjing, L. An, N. Haiying, H. Lixian, Q. Luo, and C. Yan, "Churn Analysis of Online Social Network Users Using Data Mining Techniques", *Preced. Int. multi Conf. Eng. Comput. Sci*., no. 1, pp. 14–16, 2012.
[25] W. Verbeke, D. Martens, and B. Baesens, "Social network analysis for customer churn prediction",*Appl. Soft Comput*., vol. 14, pp. 431–446, Jan. 2014. https://doi.org/10.1016/j.asoc.2013.09.017.
[26] D. Archambault, N. Hurley, and C. T. Tu (2013), "ChurnVis: Visualizing mobile telecommunications churn on a social network with attributes," *Adv. Soc. Networks Anal. Min (ASONAM)*, 2013 IEEE/ACM, pp. 894–901. https://doi.org/10.1145/2492517.2500274.
[27] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi (2008), "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 - EDBT '08*, pp. 668–677. https://doi.org/10.1145/1353343.1353424.
[28] J. David Nunez-Gonzalez, M. Grana, and B. Apolloni (2014), "Reputation features for trust prediction in social networks", Neurocomputing, vol. 166, pp. 1–7. https://doi.org/10.1016/j.neucom.2014.10.099.
[29] U. Prasad Devi and S. Madhavi (2012), "Prediction Of Churn Behavior Of Bank Customers", *Bus. Intell. J*., vol. 5, no. 1, pp. 96– 10.
[30] K. Chitra and B. Subashini (2011), "Customer Retention in Banking Sector using Predictive Data Mining Technique", *ICIT 2011 5th Int. Conf. Inf. Technol.*