International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Customer Churn Detection for insurance data using Blended Logistic Regression Decision Tree Algorithm (BLRDT)

**Nagaraju Jajam[1]*, Dr. Nagendra Panini Challa[2]**

**Submitted**: 16/10/2022 **Revised**: 22/12/2022 **Accepted**: 05/01/2023

***Abstract:*** Customer churn identification is indeed a subject in which machine learning has been used to forecast whether or not a client will exit the corporation. Predicting client turnover is a difficult issue in a variety of businesses, with the financial sector being the most well-known. Because of the constant updating of insurance plans, the client retention procedure is critical to the company's success. Predicting client turnover and developing customer retention strategies are both key research areas in the insurance market. This identifies the importance of detecting churn rate in the insurance industry. The primary objective of this research is to forecast consumer habits and to distinguish between churners and non-churners during an earlier phase. We propose a Blended Logistic Regression Decision Tree (BLRDT) framework for churn detection. Initially, the preprocessing of the insurance dataset is done by employing the Z-score technique. Data splitting is done to separate the standardized dataset into training and testing sets. The churn prediction is done using the BLRDT algorithm. The evaluation was done by using different measurements like accuracy, precision, recall, and f1-score and the outcomes are depicted by employing the MATLAB environment. The research also discovered that the presented scheme seems to be a realistic and slightly superior strategy for the insurance sector to estimate client turnover than the findings obtained using existing approaches.

***Keywords:*** *Customer churns detection, preprocessing, Z-score technique, and Blended Logistic Regression Decision Tree (BLRDT) algorithm.*

## 1. Introduction

The financial industry is one of the largest and most prominent industries that employs data forecasting technologies. Insurers' massive reliance on data and an ever-increasing customer base is a major factor in this theory. Scholars and businesses alike find this a daunting idea. Clients now have the option to transfer insurance providers, which means the churn probability projection in the insurance industry isn't meeting their demands. As a consequence, churn predictions using ML techniques are addressed. Because of its size and complexity, the financial industry undergoes rapid transformations in reaction to global economic shifts. An interesting study subject for students interested in solving real-world challenges is the development of a trustworthy and honest customer relationship management system (CRM). Insurance churn prediction is an important research area for fault detection and recognition operations. Customer churners are on the rise because of technology that can't accurately predict client turnover. Ullah, I et al. (2019) [1].According to marketing experts, businesses lose 26 percent of their clients each year due to inefficiency and the inability to search for a specific item.Acquiring a new client, on the other hand, is able than keeping an old one. Churn, defined as a consumer's willingness to discontinue a contract and go to another organization, has become a big worry in several industries. The monthly ratio for American mobile networks is 3% to 4%, which represents a significant cost for corporations that pay 450 to 550 dollors to recruit a single client who generates around 50 dollors in monthly income.Businesses are finally recognising the importance of service preservation. According to one study, "the top six American telecom companies would still have gained 207 million dollars if they could have retained an additional 5% of consumers who were ready to offer but changed contracts in the previous year."The corporation's main selling task in the coming decades is to reduce churn rates by detecting customers who are likely to depart and then making the required efforts to keep them. The first stage is to anticipate the possibility of churn at the consumer level. There are 2 significant features of the Churn Forecasting model.

- The first is that the number of churned clients (critical instances) is quite low (2% of all samples).
- The second factor is precision. As a result, increasing monthly forecasting predictive performance by 1% could result in a $54 million increase in annual revenue for a service with 1.5 million customers. Jamjoom, A.A., et al. (2021) [2].

*1* Research Scholar, School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh-522237, India. E-mail(s): nagaraju.jajam@vitap.ac.in*
*2Assistant Professor, School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, - 522237 India E-mail(s): nagendra.challa@vitap.ac.in*

Churn forecasting is a classification issue that distinguishes between regular and churn clients. The conventional classifier determines the smallest distance between two categories of data. There were multiple major applications in this area. However, the insurance churn forecast statistics are unique in that the regular database is significantly bigger than that of the irregular database. As a result, the typical classifier fails miserably at our assignment. De Caigny, A., et al. (2018) [3]. Our job is to predict what customers will do and to figure out who will and who won't be customers at an early stage. This paper presents a Blended Logistic Regression Decision Tree (BLRDT) framework for churn detection. Firstly, the preprocessing of the insurance dataset is done by employing the Z-score technique. Data splitting is done to separate the standardised dataset into training and testing sets. The churn prediction is done using the BLRDT algorithm. The evaluation was carried out by using different metrics, and the outcomes are depicted by employing the MATLAB environment. The further part of this research is: topic-II justifies the related works; topic-III explains the proposed approach; topic-IV depicts result and discussion; topic-V summarizes the complete work.

## 2. Related works

According to He, Y., et al. (2020) [4], the author is trying to find the best machine learning model for forecasting client abandonment. The data collection contains information on demographics, purchasing habits, and the macroeconomic climate as a whole. For the purpose of gaining insight into how policy duration and coverage types affect a target variable whether or not consumers renew their policies exploratory research is performed on critical aspects. Feature dimension reduction and the isolation of relevant features for use with a group of potential ML models are two of the more challenging parts of dealing with a big dataset. The ML models with the greatest Area Under the Curve (AUC) performance are the Extremely Randomized Trees Classifier and the Gradient Boosting Model.

Churn prediction models developed by the author in Ahmad, A.K., et al. (2019) [5] may assist telecom providers in analysing which clients are most likely to exit. A new method for feature engineering and selection was developed using machine learning techniques on a huge data platform. For the purpose of evaluating the model's performance, the AUC standard measurement is used. Additionally, the prediction model incorporates aspects of social network analysis (SNA) extracted from the customer's social network. The model was built and tested in the Spark environment using large amounts of raw data given by SyriaTel, a telecommunications company. Customers' data over a nine-month period was included in the dataset used by SyriaTel for system training, testing, and evaluation. It utilised algorithms like decision trees, random forests, gradient boosted machine trees (GBM) etc.A JIT method for CCP was suggested by the author of Amin, A. et al. (2018) [6]. But in order to train the classifier, the JIT approach needs prior data. For this reason, we created a JIT-CCP model based on the notion of using data from two different companies, one as a training set and the other as a goal, to fill the void in our research. There must be a rigorous modification process before data from different companies may be used to support JIT-JIT CCP's definitions. The study's objective is to provide an empirical assessment of the proposed JIT-CCP model using and without cutting-edge data transformation technologies. With Naive Bayes as their primary classifier, they experiment with publicly accessible benchmark data sets. In De la Llave, M., et al. (2019) [7], the author explains that the LLM has a segmentation step and a prediction step. In the first step, decision criteria are utilised to define customer groups, and a model is designed for every branch of the tree in the second stage. The novel hybrid approach can be matched to decision trees, random forests, and logistic model trees in terms of prediction performance and comprehension. When it comes to evaluating LLM's predictive power, the top decile lift (TDL) and area under the receiver operating characteristic GBDT curve come to mind, both of which show that it is superior to LLM's building blocks, logistic regression and decision trees, and on par with more advanced evolutionary algorithms, random forests, and logistic model trees.

In Wang, Q.F., et al. (2019) [8], the author mentions that, based on their search ad activity, customers may be predicted to be future churners using an ensemble model of gradient-boosting decision tree (). For the GBDT, they use two kinds of features: dynamic and static. It's important to keep in mind that for dynamic features, we take into account a long-term sequence of consumer behaviours (e.g. impressions, clicks). Customers' preferences are taken into account while developing static features. Using a huge amount of client information from the Bing Ads system, they assessed the prediction performance and found that static and dynamic characteristics work well together, and that integrating all of the variables yielded the best AUC (area under the ROC curve) on the test set.Our results were compared to those of other churn prediction algorithms using actual information captured from a partner firm in Al-Mashraie, M., et al. (2019) [9]. Prediction models like logistic regression, SVM, random forest, etc. Other than that, the PPM design was also used to look into the effects of certain things on customer turnover behaviour from the perspectives of push and pull, as well as mooring and anchoring the relationship. A partial least squares (PLS) regression was employed in the PPM assessment. Those who churn as well as those who don't churn are also being observed and analysed.According to Zhao, M., et al. (2021) [10], the author develops a customer churn prediction model based upon big data from high-value client operations in the telecom sector, successfully detects prospective churned consumers, and then suggests targeted win-back techniques based on the empirical study findings. They investigate the patterns and factors of client turnover using data mining techniques and give solutions to questions like how customer churn happens, the elements that influence customer churn, and how firms regain churned clients.

The author of JB Raja et al. (2020) [11] proposes a variety of feature selection tactics to help boost the accuracy of the built-in churn prediction design. The most crucial stage in improving accuracy is feature selection, which entails removing insignificant elements and emphasising those that have an impact. A decision tree was used as a classifier. As assessment metrics for the model, the TP score, FP score, accuracy, ROC, and precision were used.The author's major purpose in Abubakr et al. (2020) [12] is to use a large machine learning data platform to examine customer attrition prediction in the telecom business. The chance of a customer churning was calculated using machine learning methods. This research employs logistic regression and KNN with huge datasets to forecast customer churn in the telecom business.

The chance of churn as a function of a collection of parameters or customer attributes is often assessed using logistic regression. The K-Nearest Neighbor was also utilised to assess whether or not a customer churns depending upon the closeness of their feature to clients in every class.According to fridrich, m., et al (2019) [13], a customer churn prediction model was subjected to the impacts of several explanatory variable selection methods. The filter and wrapper approaches to parameter selection were assessed, and the novel idea of balanced clustering enhances the machine learning pipeline's runtime. Based on their simplicity of use and interpretability (LOGIT, CIT), as well as their complexity and shown performance on a given dataset, classification learners are selected (RF, RBF-SVM). Even when used in combination with a learner efficient at incorporating characteristic selection, ceratinparameter selection may not necessarily result in better achievement.

In Calzada-Infante,et al. (2020) [14] propose an innovative approach to extracting the dynamic importance of every client, utilising social network test tools and a binary classification algorithm. The dynamic relevance of every client has been derived by using multiple centrality measures across temporal charts to depict the interactions among customers as well as to eliminate churners and non-churners. A temporal chart is constructed by using the call detail data of telco clients in order to build these connections.Depending upon the aforesaid notion of the classifier's confidence estimate utilising the distance element, a unique CCP technique is provided in Amin, A., et al. (2019) [15]. The database was separated into areas based on the distance element and then into 2 types: (i) information with maximum confidence and (ii) information with minimum certainty in forecasting consumers who show churn and non-churn behaviour. Accuracy, precision, and recall are all measurements of how well a classifier predicts whether a customer will churn or stay loyal to a service provider. That's not true, though. The distance factor is very important to the classifier's level of certainty. The classifier did better in the area with the highest distance element value.Research shows that roles are being reversed, and firms have pledged to demonstrate customer loyalty using CRM that eliminates turnover before it occurs. Schena, F. et al. (2016) [16]. When a company knows how to read data and use technology, it can save money on IT costs by predicting what customers will do before they leave, which helps them keep their customers. Classifier The author of Stripling, E., et al. (2018) [17] uses a genetic algorithm to maximise EMPC during training, where ProfLogit's internal model structure mimics one of the most basic types of logistic models: the lasso-regularized one. Using the predicted profit-maximizing fraction, we also offer threshold-independent recall and accuracy measurements. To meet the corporate need for profit maximisation, we propose a strategy for creating lucrative churn models for retention initiatives. Overall, ProfLogit shows the top EMPC performance and profit-based accuracy and recall scores in a benchmark analysis with nine real-world examples.

By introducing the MPU measure in Devriendt, F., et al. (2021) [18], the author proposes a new metric for evaluating work regarding the greatest potential benefit that may be attained by using an uplift design. Extending the maximum profit metric created for analysing customer churn uplift designs is the purpose of this measurement. The liftup curvature and liftup measurement in assessing uplift designs are analogues of the lift curvature and lift measurement, which were widely employed to assess predictive frameworks while introducing the MPU measure. It was then used to look at and compare the models' performance in a financial sector case study. Measures of customer churn anticipation and uplift model performance were used to do this.According to Vo, N.N., et al. (2021) [19], an unstructured data-based customer churn prediction model is proposed by the author. Extensive tests were done using data from a large contact centre dataset consisting of 2 million calls from over two hundred thousand clients. Employing interpretable machine learning with personality characteristics and client segmentation, the findings reveal that our model properly predicts client churn risks. In Pustokhina et al. (2021), [20] CCPBI-TAMO is an algorithm that uses text analytics and metaheuristics to create a dynamic CCP approach for industry intelligence. For the feature selection, the chaotic pigeon-inspired optimization (CPIO) based feature selection (FS) approach, which minimises computing complexity, is used. Additional classification methods include long-term memory (LSTM) and auto encoder (SAE) models, which were utilised to produce characteristic-reduced information. The SAE-LSTM design incorporates SAE's capacity to recognise small characteristics into the LSTM model's ability to classify data. As a final step, the hyperparameter tuning procedure known as "sunflower optimization" (SFO) takes place.Particle swarm optimization (PSO) with feature selection as a preprocessing methodology, PSO plus simulated annealing, as well as PSO including both feature selection and simulated annealing, are the 3 phases of PSO recommended by the author of Vijaya et al. (2019) J [21] in churn forecasting. To evaluate their prediction levels and performance elements, the suggested classifiers have related to decision trees, nave bayes, K-nearest neighbour, and three hybrid techniques, etc.

Jeyakarthic M, et al. (2020) [22] In the cloud computing context, this research provides a revolutionary CCP model based on ML algorithms. Data gathering, preprocessing, and adaptive gain with back propagation neural networks (AGBPNN) are utilised in the suggested CCP model. A client database would be collected initially using different IoT devices such as computers, smart phones, wearables, and so on. The data gathered by the IoT applications was transmitted to a cloud data server (CDS). Following that, preprocessing occurs, during which the dataset's missing values are efficiently computed. After that, the P-AGBPNN framework is run on the cloud to know if the client is a churner or not. The AG-BPNN design's outcomes were analysed by a benchmark database from the telecommunications industry. In De Bock, K.W. et al. (2021) [23], rule ensembles and their extension, spline-rule ensembles, are introduced as a potential family of classification techniques for predicting customer turnover. The flexibility of a tree-based ensemble classifier is combined with the ease of regression analysis in spline-rule ensembles. They do, however, overlook the interrelationships between potentially conflicting model components, which may lead to needless model complexity and damage model interpretability. Spline-rule ensembles with lasso regularisation, a unique algorithmic extension, are shown to make it easier for people to understand what they're seeing.For instance, in Li, Y., et al. (20210 [24], at every step of the supply chain, newcomers are vying for the attention of China's national broadcast service providers. Despite the benefits of the past, the customer turnover rate has been rising in recent years. It is important for national

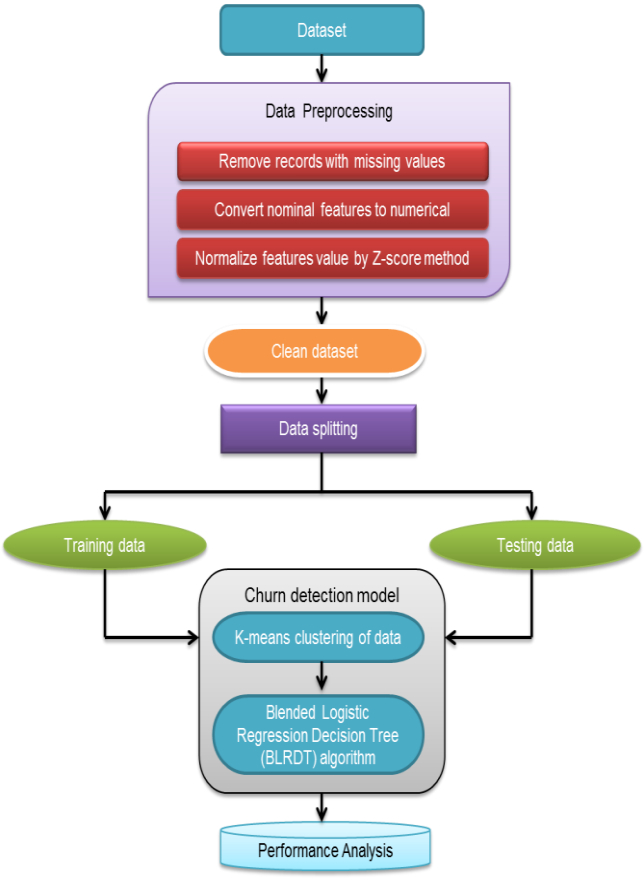
broadcast service providers to know their customers' preferences and anticipate client attrition before the competitor does. According to a positivist perspective, customer turnover is linked to the frequency, quantity, and mode of consumption as well as the method of payment. In terms of consumer viewing intensity and customer turnover, watching preference has just a minor impact. Interviews with users were conducted to find out how these factors interacted and what techniques they used to stay engaged. In this study, businesses that are similar to big, traditional businesses are discussed.In Sivasankar, E. et al. (2019) [25], you will find In this study (PPFCM-ANN), hybrid probabilistic C-means clustering (PPFCM) and artificial neural networks are used to predict customer turnover in this study (PPFCM-ANN). A clustering component is proposed related to the PPFCM, and a churn anticipation is proposed using an artificial neural network (ANN) in this research. A probabilistic C-means clustering technique could be employed in the clustering module to group the input dataset. The grouped dataset was employed in the artificial neural network, and this hybrid architecture was utilised in the churn prediction design. For each grouped test set, ANN classifiers with the lowest distance or similarity scores are chosen from among those that have been trained using those data. In the end, the output score value is used to estimate client attrition. In Vijaya, J. et al. (2018) [26], a strategy for predicting client turnover in the telecom sector was presented using RST. Using ensemble-classification techniques like bagging, boosting, and random subspace, the chosen features are then classified. There are three experiments in this study that use the Duke University-Churn prediction dataset in assessment. This model is better than any other single model at classifying people correctly than any other model. This is based on criteria like true churn, false churn, specificity, and accuracy.The risk of churn is precisely calculated and correlations between risk and customer behaviour are identified in Routh, P., et al. (2021) [27] using a competing risk technique inside a random survival forest methodology. Existing techniques depend on certain functional forms to explain risk-behavior interactions; the proposed model does not. It also does not contain distributional assumptions, which are both practical constraints. The method's effectiveness is examined using data from a hospitality-industry membership-based organisation, where clients encounter two competitive churning events.

As stated in the Pamina, J., et al. (2019) [28] passage, as a result of this research, telecom carriers can detect consumers who are going to churn more easily. Emphasis is placed on the recall assessment metric, which provides an immediate solution to the business challenge at hand. Churn analysis is all about using customer information to identify those who are most likely to exit the job in the near future. At least 15 separate machine learning algorithms are used, each with its own set of parameters. According to Ahn, J., et al. (2019) [29], the author compiled solutions to churn utilised in the sectors of industry organisation and sales, as well as IT, telecommunications, media, insurance, and psychology. Accordingly, we categorised and illustrated churn loss, feature engineering, and predictive designs. Researchers interested in the service sector may utilise our study to integrate fragmented churn data from industry and academia to determine the best solution for churn and the accompanying designs that are most appropriate.InGerber, G., et al (2020) [30], the author intended to adjust for the effect of random censoring. They investigate both scenarios in which the censoring is independent of the variables and scenarios in which it is not.They use simulated and actual data analysis to compare our technique to other established approaches that are applicable in our context. We demonstrate that the solution was more aggressive than the given issue, which involves a quadratic mistake. On the Internet, you may find further information and resources on this topic.

In Amin, A. et al. (2020) [31], the author explored information from a different firm in the context of JIT to solve CCP issues in the telecommunications industry. Researchers used publicly accessible information from 2 telecom firms to test the suggested model's performance. To see how well the predictive model works in the JIT-CCP context, you can look at cross-company datasets. In this case, the JIT-CCP approach is better than individual classifiers or a homogeneous ensemble-based strategy.In Scriney, M. et al. (2020) [32], the author outlined a method for producing some of the missing CLV parameters. In this case, it has a new way of filling in the gaps in the data. In Morik, K. et al. (2004) [33], the author presented the results of a comprehensive method of knowledge discovery based on insurance policy information. The TF/IDF representations from data acquisition are used to compile the time-related aspects of the data set. These additional characteristics have been shown to improve accuracy, precision, and recall in experiments. A heuristic is used to figure out how much the feature space grows or shrinks as a result of the TF/IDF transformation. In Rai, S., et al. (2020) [34], the author built the call tree classification model, analysed it, and compared it to the logistic regression model in terms of performance indicators.In Mishachandar, B et al. (2018) [35], the author proposed an innovative method for churn prediction by integrating machine learning models with Big Data Analytics technologies. Predicting client churning behaviour in advance is the goal of the focused, proactive retention method. In order to better understand the risk of client churn, telecom businesses would benefit from this recommended research.In Jain, H. et al. (2020) [36], the author constructed a framework, and 2 machine-learning approaches, logistic regression and logit boost, have been used to analyse customer attrition. Real information from an American firm was used for the experiment in the WEKA Machine Learning tool. The results were presented in a variety of waysIn Devriendt, F., et al. (2021) [37], the author evaluated effectiveness by utilising the maximum profit uplift (MPU) measure, which assesses the highest profit that can be gained by implementing an uplift strategy. As an extension of the greater profit metric, this measure was created to evaluate customer churn uplift approaches. The liftup curve and measures for assessing uplift models are analogues of the lift curve and measures that are widely used to assess predictive models while presenting the MPU metric. Using uplift models, researchers have found that efforts to keep people from leaving are more profitable because they go above and beyond the predicted models.

In Jin, H., et al. (2022) [38], the author reacted to financial incentives by switching plans or cancelling their connection with the mobile network operator in [38]. Simulated policy simulations reveal that the operator should urge customers to limit their expenditures given the trade-off between profits from overspending and customer turnover encountered by the operator. As we demonstrate in our research, the consequences of switching choices on future expenditures and the risk of churn vary between upward and downward switchers as well. With regard to reducing

future expenditures, customers who move down are more likely to do so than those who move up.In Dalli, A., et al. (2022) [39], the author examined numerous published publications that employed machine learning (ML) approaches to forecast a churn in the telecommunications industry. Deep learning approaches have shown substantial predictive power. Adadelta, Adam, AdaGrad, and AdaMax algorithms all showed improved results for RemsProp.InShabankareh, M.J., et al. (2021) [40], the author suggested a layered data mining method to deliver an efficient early churn detection solution. Research shows that combining SVM with the chi-square automatic interaction detection (CHAID) decision tree is the best way to get the most accurate outcomes. The findings reveal that the suggested churn anticipation system has the correct accuracy. Customer churn detection was also enhanced because of the stacking technique.




## 2.1 Problem statement

Insurance companies competed fiercely for the business of people all over the world.Due to the obvious recent increase in the highly competitive environment of health insurance businesses, clients are migrating. It's unclear if there is evidence of switching behaviour and which customers go to a competitor. When there are so many different pieces of information recorded from millions of clients, it's tough to study and comprehend the causes of a customer's decision to switch insurance carriers. In an industry where customer retention is equally important, with the latter being the more costly method, insurance companies depend on the information to evaluate client behaviour to minimise loss. Insurance firms may design techniques to prevent customers from actually transferring if they know whether or not they are likely to do so ahead of time. The ability to accurately estimate future attrition rates is crucial since it helps the organisation comprehend future earnings. Churn level predictions may also help your firm identify and improve areas where customer service is lacking. As a result, the blended logistic regression decision tree technique is used to investigate machine learning-based customer churn detection for insurance data in this research.

## 3. Proposed work

In this section, machine learning based churn detection for insurance data using blended logistic regression algorithm is discussed. Dataset is preprocessed using z-score normalization method and the processed data is then split into training and testing data. By using training and testing data, churn detection is identified using the proposed blended logistic regression decision tree algorithm (BLRDT). And its performance are analyzed using accuracy, precision, recall and f1-score. Figure 1 represent overall flow of our research.

### A. Dataset collection.

Our client is an insurance firm that has given health insurance information to its consumers. They need your help in developing a model to predict whether the previous year's policyholders (customers) will be interested in the firm's vehicle insurance. An insurance policy is a contract in which a corporation agrees to pay a certain premium in exchange for the assurance of reimbursement for a specific loss, destruction, disease, or mortality. A premium is a set amount of money that a consumer must pay to an insurance firm in exchange for this assurance. Vehicle insurance is similar to medical insurance in that the client pays an annual fee to the insurance provider business so that, in the event of an unfortunate accident caused by the vehicle, the insurance provider company will compensate the customer. Developing a method to forecast if a client is interested in vehicle insurance is incredibly beneficial to the organization because it allows it to design its communications strategy to reach out to those consumers and utilize its business strategy and income appropriately. For the ability to forecast whether a client would be keen on vehicle insurance, data about demographics (gender, age, region code method), vehicles (vehicle age, malfunction), policies (premium, sourcing channel), and other factors such as demographics (gender, age, region code type), vehicles (vehicle age, damage), and policies (premium, sourcing channel) is needed. In this study, the main goal is to make an effective churn prediction system and to study the information visualization findings as thoroughly as possible. The database has been collected from the Kaggle open data website (https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction? select=test.csv). The total size of the database is 45,211. We divided this data set into two different sets. The training set has 33,908 data points, and the test set has 11,303 data points.

### B. Data pre-processing

Initial stage is to plan the dataset to realize proper input. Raw data is retrieved among all datasets in this phase, after which it is handled in a following stage. The major objective of this method is to detect any possible issues or inaccuracies in the information that may have arisen during the information gathering

process and attempt to rectify it prior moving on with the next phase.

- **Remove records with missing values**

As previously said, pre-processing stage is a stage of data separation process that aims to eliminate missing values before investigators explore and extract hidden information from information. In actual statistics, there have been frequently missing values. Data redundancy or a missing important information may result in missing data. To put it differently, the existence of missing data limits the formation of learning methods, and as a consequence, information hidden in statistics is not gained. There have been seven options for dealing with missing values in the database:

- **Deleting Rows with missing values**

Eliminate the rows/columns with zeros to deal with missing values. When greater than 50% of the rows in a column become empty, then column could be erasedcompletely. The rows with zeros in one or more columns could also were erased.

- **Other Imputation Method**

Alternative imputation techniques might be a little more suitable to restore missing values relying on the scope of the information or data category. As instance, if the information parameter has longitudinal activity, it mayarrange sense to fill the missing value with the most fresh accurate measurement. Last observation carried forward (LOCF) tool can be used for it. For a missing value in a time-series information parameter, it makes sense to perform the variable's interpolation before and after a timestamp.

- **Using Algorithms that support missing values**

Missing values are not sustained by all machine learning tools, but certain ML methodologies are adaptable to missing values in the database. Whenever a value is missing, the k-NN method may leave out a column from a distance metric. While making a anticipation, Naive Bayes could also considered into account missing values. If a data consists null or missing values, some techniques could be applied.

- **Imputation using Deep Learning Library — Datawig**

Considering categorical, continuous, and non-numerical characteristics, this technique excels. Datawig is a toolkit that uses Deep Neural Networks can develop machine learning methods and replace missing values in datagrams. With all other columns as sources, Datawig may fit a restoration strategy for every column with missing values in a data set.

- **Convert multi value feature to nominal feature**

Converting multi-value characteristics to numerical features would be another stage in pre - processing phase. While multi-value characteristics could be useful in some methods, neural networks don't really accept them. As a result, data preprocessing was regarded as a crucial phase in the data preparation phase.

- **Normalize feature value by z-score method**

The technique of standardizing each value in a database so that the mean of all of the values equals 0 as well as the standard deviation equals 1 is known as Z-score standardization. We used the Z-transform process of information standardization for the purpose of this research and to conduct the normalisation phase. Furthermore, it has advocated that in order to ease the modelling process, sampling approaches be used, as massive data evaluation can be costly and time consuming. To conduct a z-score normalisation on each value in a database, we use the following expressi

$$New\ value = \frac{(x-\mu)}{\sigma} \quad (1)$$

where x=original value, $\mu$=mean of data, and $\sigma$ = standard deviation.Data cleaning seems to be essential for reducing the data collection's dimensions. As the dimension grows, so does the amount of time & computing power needed.

### C. Clean dataset

After the preprocessing stage, the quality of data is improved by removing the errors and inconsistency . Clean dataset refers to the data with free of errors. The process of occuring clean dataset is also known as scrubbing. Data quality problems occur because of misspellings during entry of information, or any other invalid information. Basically, “garbage” information is transformed into clean information. “garbage data” does not produce the accurate and good outcomes. So it becomes very important to handle this data.

### D. Data splitting

Data dividing is indeed a method of dividing a database into at least 2 subgroups, referred to as 'training' (or 'calibration') and 'testing' (or 'prediction'). This phase is normally performed after the samples' spectra were adjusted for distortion or unwanted variation during pre-processing. The information that would be input into the model would be stored in the training dataset. The data used to evaluate the trained & verified strategy is contained in the testing dataset. It indicates how effective our overall model is and how frequently it would be to anticipate something that is incorrect.As mentioned above in our study, training set consists of 33,908 data, whereas testing set contains 11,303 data

### E. K-Means Clustering Algorithm

If you have a set of n observations or data represented with {x1,x2,…,xn}, and each data includes d features and no labels, k-mean clustering provides a simple method to group the whole data set into k clusters {c1,c2,…,ck} , where data in the same cluster are more similar to each other. Each data (xi ) is labeled with the cluster j(1…k), where the distance of the data point i to the centroid of cluster j is minimum compare to all other clusters.Different methods for initializing this unsupervised clustering algorithm

exist, and this initialization is important for reducing the complexity of the algorithm (the number of iterations required) and the results of clustering (local optimum). Typically, if the initial centroids are closer to the final centroid, you can achieve better clustering results and lower complexity. This application note does not cover initialization methods.

### F. Churn Detection Model.

The prediction model for the churn detection technique would be built using a blended logistic regression decision tree technique, that would be used to assess the system's effectiveness. A decision tree is a visual representation of an algorithm with just conditional control structures. They're often used in operations research, notably in decision support, to assist discover the best method for achieving a target, but they're also a popular machine learning approach. CART (Classification and Regression Trees) divides the feature space into nonoverlapping sections in a recursive manner. A classification tree is created to forecast the evolution of a dependent categorical variable. For evaluate the goodness of fit more precisely, CART involves both testing with test insurance data & cross-validation. CART could use the same parameters in various regions of the tree at the same time. Complex interdependencies among sets of variables could be discovered using this feature. To choose the input set of insurance data, CART could be used in association with other forecasting models. The insurance dataset's most essential parameters were chosen for decision tree visualisation, and predictions are created as a result. The maximum value feature has been used to divide the insurance information in the decision tree algorithm, and the information gain has been determined for every characteristic. This is repeated till all the characteristics have been used, and then the characteristics who don't provide sufficient information are pruned from the tree to have an enhanced tree for the highest suitable assessment.Decision trees could provide a simple visual overview of the possible information through which guidelines could be produced and schemes for customer retention could be built, and logistic regression enables us to know the degree to which each characteristic influences the decision to churn. Both logistic regression and decision trees offer benefits and weaknesses that are relatively complementary. The former employs a basic (linear) statistical model, and the fitting process seems to be very steady, leading to decreased variation but potentially major bias. The latter, on the other hand, has a low bias but a large variation since it examines a smaller range of models, which allows it to detect nonlinear patterns from the data but also makes it less steady and susceptible to overfitting. To resolve those weaknesses, in this study, we proposed a blended logistic regression decision tree (BLRDT) algorithm. This model is obtained by combining the logistic regression and decision tree algorithms on the feasible insurance database.

The dependent parameter is transformed into a BLRDT parameter using maximum likelihood estimation within BLRDT. This estimates the parameters for the linear function using the linear regression model for assess the value of the dependent parameter.As depicted in (1) $\alpha, q_1, q_2, \ldots q_n$ are the variables to be assessed employing the training dataset& the equation is utilized to detect Tthatdepicts the dependent parameter value whenvalues of characteristics x1,x2,…xn werespecifie

$$T = p + q_1y_1 + q_2y_2 + \cdots + q_ny_n \qquad (2)$$

It's employed whenever the outcome becomes binary, such as yes or no, Zero or 1, and so on. Because the result of expression (2) is indeed a real-valued numeric, it must be translated into a form that can be used to make the forecast. Thus, to turn the result of linear regression into a probability level, logic or the sigmoid function was applied, as well as its expression is as shown in:

$$S = \frac{1}{1+e^{-(p+q_1y_1+q_2y_2+\cdots+q_ny_n)}} \qquad (3)$$

$$0 \leq S \leq 1$$
$$-\infty < T < +\infty$$

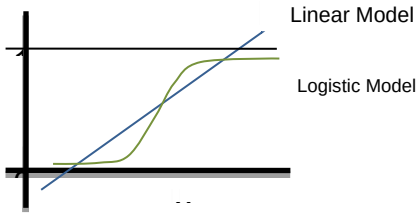



**Fig. 2** Steepness of the curve

In the BLRDTconstant (p) shifts the curvature left-side& right-sideas well as the slope ($q_1$) specifies the steepness of the curvature in Figure 1. The BLRDT equation could be specified in terms of an odds ratio with a basic transformation.

$$\frac{S}{S-1} = \exp(p + q_1y_1 + q_2y_2 + \cdots + q_ny_n) \qquad (4)$$

Furthermore, we may define the expression in terms of log-odds (logit) that would be a linear function of the predictors, by considering the natural log of both sides. The component ($q_1$) indicates how much the logit (log-odds) varies when y has been changed through one unit

$$n\left(\frac{s}{s-1}\right) = p + q_1y_1 + q_2y_2 + \cdots + q_ny_n \qquad (5)$$

BLRDT could manage whatever quantity of quantitative and/or categorical variables, as previously stated. Because S is a value linking 0 and 1, it can be utilizd to estimate the chance of a desired outcome. For instance, if the result is 0.8, it signifies that there have been 80% odds of having the result as 1, and that can indeed be fairly forecasted that the result will be 1 for the specified combination of input qualities. The likelihood of the negative scenario, or 0 in this situation, can be computed by subtracting 1 from the probability of the positive argument. For the previous case, it will be 0.2, which is significantly smaller, and hence the forecast will be 1. Usually, a threshold of 0.5 is utilized to find out which prediction must be finished. Any value greater than 0.5 is measured positive, whereaseverything below is measured negative. To enhance the performance of the churn prediction model

$$a(i;\theta,j) = \theta_1i_1 + \theta_2i_2 + \cdots \ldots + \theta_si_s + j \ldots \ldots \qquad (6)$$

$a(i;\theta,j)$ determined by the value of $\theta_1, \theta_2$ ……$\theta_k$ and factor $j$ , denotes the co efficient, $i_1, i_2, i_3$…….$i_s$ and denotes the feature vector. Equation 6 can be denoted in vector form as f

$$a(i;\theta,j) = \theta^P i + j \quad (7)$$

The sigmoid function is used to map anticipated values between 0 and 1, as illustrated in the equation.

$a(i;\theta,j) = r(\theta^P i + j), where$

$$r(g) = \frac{1}{1+\exp(-g)} \quad (8)$$

The following equation may be used to create a mathematical churn prediction model:

$$a\left(i^{(1)};\theta,b\right) \approx c^{(1)} \ldots\ldots \quad (9)$$

$$a\left(i^{(2)};\theta,j\right) \approx c^{(2)} \ldots\ldots \quad (10)$$

……

……..

---

$$a\left(i^{(n)};\theta,j\right) \approx c^{(n)} \ldots\ldots \quad (11)$$

$a\left(i^{(1)};\theta,b\right) \approx c^{(1)}$ Indicates the class label for 1st customer, $a\left(i^{(2)};\theta,b\right) \approx c^{(1)}$ indicates the class label for 2nd Customer, mth customer class label is depicted as $a\left(i^{(n)};\theta,j\right) \approx c^{(n)}$

The following mean square error approach is utilized to reduce the objective function to a minimum:

$$h(\theta,j) = \left(a\left(i^{(1)};\theta,j\right) - c^{(1)}\right)^2 + \left(a\left(i^{(2)};\theta,j\right) - c^{(2)}\right)^2 + \cdots\ldots..+\left(a\left(i^{(n)};\theta,j\right) - c^{(n)}\right)^2 = \sum_{l=1}^{n}\left(a\left(i^{(l)};\theta,j\right) - c^{(l)}\right)^2 \quad (12)$$

**Algorithm 1 ::** Blended logistic regression decision tree Algorithm (BLRDTA

---

**Input:** (new) data $D_{val} = \{(Xi, Yi)\}_{i=1}^{N}$

1:Use decision guidelines of model M on $D_{val}$ spanning the overall space S, resulting In subspace $S_t$depending upon a collection of terminal nodes T for which $S = \cup_{t\in T} S_t$,

$\forall t \neq t'$: $S_t \cap S_{t'} = \emptyset$

2: For i=1 to T do:

3: Apply logistic regression specific for $S_t$

4: For j=1 to $n_i$ do:

5: Estimate predictions for all $n_i$ instances in $S_t$

6: End For;

7: End For;

8: Combine predictions

**Output**: one prediction for every instance in S

## 4. Result and Discussion

In our study, Matlab simulation tool is used to run the existing and proposed approach. Specifications such as a) Accuracy, b) Precision, c) Recall, and d) F1-score e)y-hat are used to validate the suggested approach's behavior. This evaluation will take into account four factors: $t_p$denotes True Positive, $t_n$ denotes True Negative, $f_p$ denotes False Positive, and $f_n$ denotes False Negative. Table 1 represents the comparison of different parameters.

- $t_p$denotes that the data is normal, and it turned out to be exactly that
- $t_n$denotes that the data is expected to be churn, and it is really

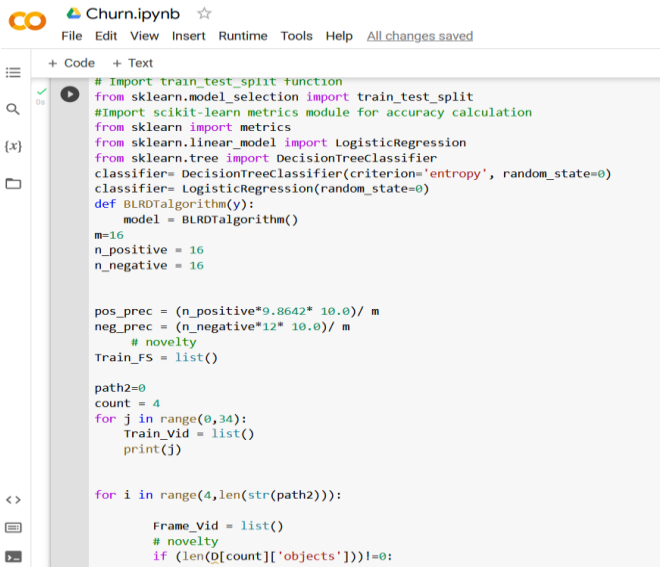

churn.

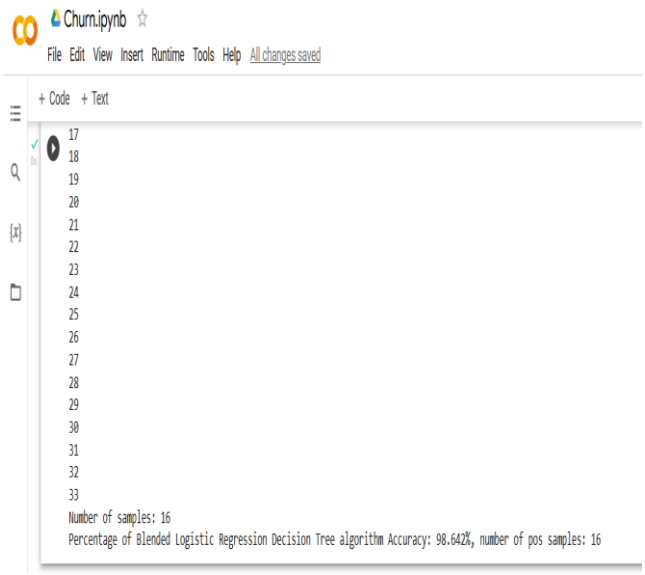
- $f_p$ denotes that the data is expected to be churn, yet it is a normal data.
- $f_n$denotes that the data is expected to appear normal, however it is churn.
- The proposed method is also compared to a few different ways for determining these values.

The confusion matrix of the testing data for the churn prediction is given as follows:

| N=11,303 | Predicted : NO | Predicted : YES |
|---|---|---|
| **Actual : NO** | 250 | 50 |
| **Actual : YES** | 3 | 11,000 |

### Code for BLRDT model:



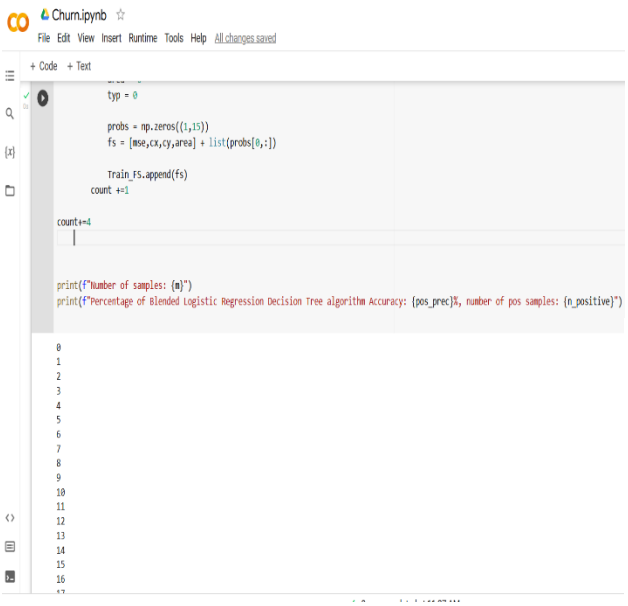



| Algorithm | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Bagging | 90 | 89 | 90 | 89 |
| Extra tree | 90.23 | 68.46 | 65 | 71 |
| XGBoost | 89 | 70 | 57 | 59 |
| BLRDT[Proposed] | 98.642 | 97 | 97 | 97.2 |

**Table 1** Comparison table

## Accuracy:

It determines the number of data that are successfully classified. It decides how closely the outcomes match the initial outcome




$$Accuracy = \frac{t_p+t_n}{t_p+t_n+f_p+f_n} \quad (13)$$

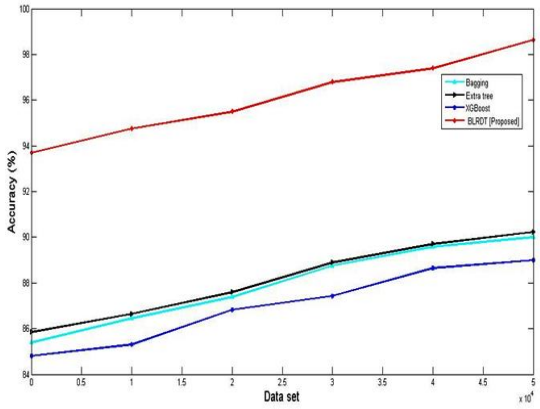



**Fig 3** Comparison of Accuracy

Figure 3 compares the accuracy of the existing methods such as Bagging (), Extra tree (), XGBoost with the proposed methodology (). The graph clearly depicts that the suggested methodology is far better than the existing methods. Table 2 indicates the accuracy of different techniques.

**Table 2** Accuracy of various techniques

| Dataset | Bagging | Extra tree | XGBoost | BLRDT[Proposed] |
|---|---|---|---|---|
| **1** | 85.75 | 86 | 84.75 | 95 |
| **2** | 87.5 | 87.75 | 85.5 | 96 |
| **3** | 88 | 88.25 | 86.5 | 97.5 |
| **4** | 88.5 | 88.75 | 87.85 | 98 |
| **5** | 90 | 90.23 | 89 | 98.642 |

## Precision

Figure 4 Precision comparison using existing and new methods

It calculates how accurate the proposed technique's behavior is, by analyzing the actual true positives from the anticipated ones. It determines how accurate the proposed technique's behavior is by separating genuine positives from false positives.

$$Precision = \frac{t_p}{t_p+f_p} \quad (14)$$

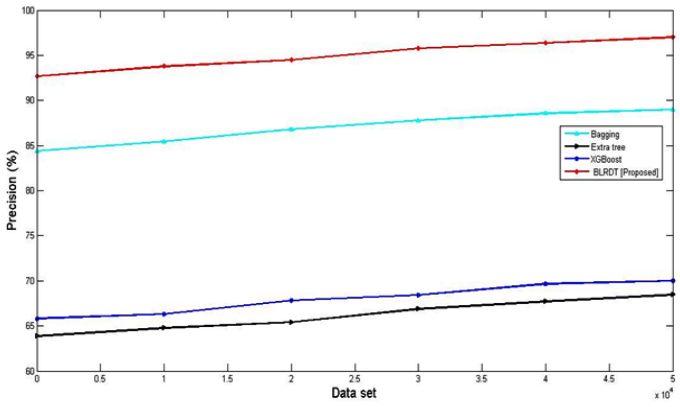


Figure 4 compares the precision of as Bagging (), Extra tree (), XGBoost with the proposed methodology (). The graph clearly shows that the new approach is more exact than previous methods. Table 3 shows the precision of different techniques.

## Recall

**Table 3** Precision of various techniques

| Dataset | Bagging | Extra tree | XGBoost | BLRDT[Proposed] |
|---|---|---|---|---|
| **1** | 85 | 65.5 | 66.25 | 92.5 |
| **2** | 86.5 | 65.75 | 67.5 | 93.5 |
| **3** | 87.5 | 66 | 67.25 | 95 |
| **4** | 88 | 66.25 | 68 | 96 |
| **5** | 89 | 68.46 | 70 | 97 |

The recall, also known as sensitivity, is the proportion of total significant samples collected**.**

$$Recall = \frac{t_p}{t_p + f_n} \qquad (15)$$

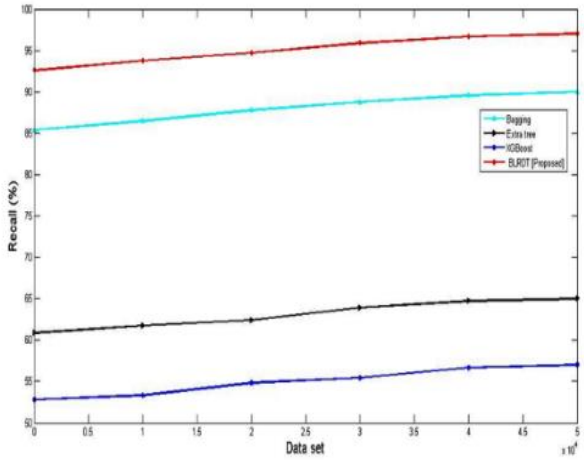



**Fig 5** Recall comparison using existing and new methods

Figure 4 compares the recall of Bagging (), Extra tree (), XGBoost with the proposed methodology (). The figure clearly shows that the new strategy is superior to the traditional systems. Table 4 shows the Recall of different techniques.

Table 4 Recall of various techniques

## F1-Score

The F1-Score is the weighted average of Precision and Recall. As a consequence, both false positives & false negatives were considered in this outcome

$$F1 - Score = \frac{2 \times precision \times recall}{precision + recall} \qquad (16)$$

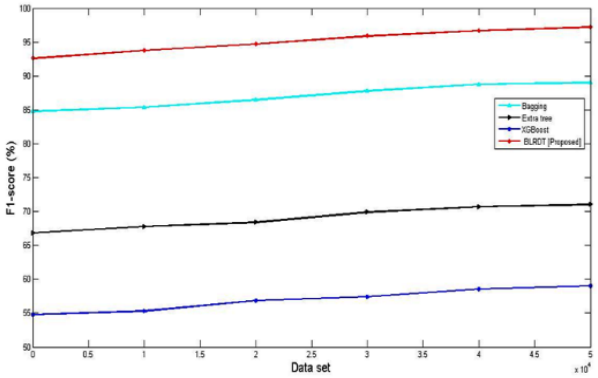



**Fig. 5** F1-Score comparison using existing and new methods

Figure 4 compares the precision of as Bagging (), Extra tree (), XGBoost with the proposed methodology (). Table 5 compares the parameters of traditional methods with the proposed method.

| Dataset | Bagging | Extra tree | XGBoost | BLRDT[Proposed] |
|---|---|---|---|---|
| 1 | 85 | 67 | 55 | 92.5 |
| 2 | 86 | 67.25 | 56 | 93.5 |
| 3 | 87 | 68 | 57 | 95 |
| 4 | 88 | 69 | 58 | 96.5 |
| 5 | 89 | 71 | 59 | 97.2 |

T

**Table 5** F1-Score of various techniques

## y-hat

The y-hat values are the assessed or forecasted results in a regression or other prediction method. It can also be considered as the average value of the response variable.

| Dataset | Bagging | Extra tree | XGBoost | BLRDT[Proposed] |
|---|---|---|---|---|
| **1** | 86.25 | 62 | 52.5 | 92.5 |
| **2** | 87 | 63.5 | 53.5 | 93.5 |
| **3** | 88 | 64 | 55 | 95 |
| **4** | 89 | 64.5 | 56 | 96 |
| **5** | 90 | 65 | 57 | 97 |

| | yhat | yhat_lower | yhat_upper |
|---|---|---|---|
| 0 | 36293.420300 | 28196.686859 | 43647.403609 |
| 1 | 37063.232217 | 29553.555299 | 44606.317588 |
| 2 | 42266.514833 | 34507.678098 | 49576.320828 |
| 3 | 44392.187651 | 36464.850347 | 51855.479108 |
| 4 | 45902.293693 | 37989.951931 | 53202.667705 |

## 5. Conclusion

Relying on a blended logistic regression decision tree, the proposed scheme shows a statistical survival evaluation technique for predicting client turnover. According to the presented model, machine learning approaches could be a promising alternative for client attrition control. The ideal churn model isn't the one with the highest statistical precision; it's the one that gives you the most insights into how to avoid churn. It would be simple to construct retention policies as well as strategies to keep clients using the obtained results, which use blended logistic regression decision tree, because these methodologies offer quickly deducible descriptions of the purposes for churning as well as a list of clients with a high possibility of churning. As previously stated, the main objective of this study was to demonstrate how innovative machine learning approaches could aid both practitioners and researchers in collecting and analysing the data more efficiently and precisely. Consumers are, without a doubt, the most valuable asset, regardless of the sector or the size of the company. From a managerial standpoint, it is critical to create an accurate method for predicting customer turnover in order to have success. As a result, businesses must always give particular attention to identifying consumers with a high chance of churn. It is able to stop the destruction of company assets by detecting churn candidates ahead. It is suggested that data on customer churn from domestic businesses be used to continue this investigation and broaden the findings. It is also suggested to simply apply specific algorithms like various forms of decision trees with interpretable findings in order to perform a precise evaluation of customer turnover factors and determine the most critical elements impacting consumer churn. This issue may reduce the accuracy of the final outcome, yet it can provide a proper understanding of customer turnover factors. This research also suggested that future experiments could combine more than two methods by bagging or stacking them together.


## Acknowledgments

The authors are grateful to Dr. Nagendra Panini Challa for the assistance on the Blended Logistic Regression Decision Tree (BLRDT) framework for churn detection algorithm

## Data Availability

No Data Availability

## Funding

Not Applicable funding sources


## Conflict of interest statement

I (we) certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

## References


[1] Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S.U. and Kim, S.W., 2019. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 7, pp.60134-60149.

[2] Jamjoom, A.A., 2021. The use of knowledge extraction in predicting customer churn in B2B. *Journal of Big Data*, 8(1), pp.1-14.

[3] De Caigny, A., Coussement, K. and De Bock, K.W., 2018. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), pp.760-772.

[4] He, Y., Xiong, Y. and Tsai, Y., 2020, April. Machine Learning Based Approaches to Predict Customer Churn for an Insurance Company. In 2020 *Systems and Information Engineering Design Symposium (SIEDS) (pp. 1-6). IEEE.*

[5] Ahmad, A.K., Jafar, A. and Aljoumaa, K., 2019. Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data, 6(1), pp.1-24.*

[6] Amin, A., Shah, B., Khattak, A.M., Baker, T. and Anwar, S., 2018, July. Just-in-time customer churn prediction: With and without data transformation. In 2018 *IEEE congress on evolutionary computation (CEC)* (pp. 1-6). IEEE.

[7] De la Llave, M.Á., López, F.A. and Angulo, A., 2019. The impact of geographical factors on churn prediction: an application to an insurance company in Madrid's urban area. *Scandinavian Actuarial Journal*, 2019(3), pp.188-203.

[8] Wang, Q.F., Xu, M. and Hussain, A., 2019. Large-scale ensemble model for customer churn prediction in search ads. *Cognitive Computation*, 11(2), pp.262-270.

[9] Al-Mashraie, M., Chung, S.H. and Jeon, H.W., 2020. Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering*, 144, p.106476.

[10] Zhao, M., Zeng, Q., Chang, M., Tong, Q. and Su, J., 2021. A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China. *Discrete Dynamics in Nature and Society*, 2021.

[11] JB Raja, G Sandhya, SS Peter, R Karthik, F Femila, 2020,Exploring effective feature selection methods for telecom churn prediction ,*Int. J. Innov. Technol. Explor. Eng.* (IJITEE) 9 (3).

[12] Abubakr, J.M.B., 2020. Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform.

[13] fridrich, m., 2019. explanatory variable selection with balanced clustering in customer churn prediction. Ad Alta: *Journal of Interdisciplinary Research*, 9(1).

[14] Calzada-Infante, L., Óskarsdóttir, M. and Baesens, B., 2020. Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts. Expert Systems with Applications, 160, p.113553.

[15] Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. and Anwar, S., 2019. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94, pp.290-301.

[16] Schena, F., 2016. Predicting customer churn in the insurance industry: A data mining case study. In Rediscovering the Essentiality of Marketing , *Springer, Cham* ,(pp. 747-751).

[17] Stripling, E., vanden Broucke, S., Antonio, K., Baesens, B. and Snoeck, M., 2018. Profit maximizing logistic model for customer churn prediction using genetic algorithms. Swarm and Evolutionary Computation, 40, pp.116-130.

[18] Devriendt, F., Berrevoets, J. and Verbeke, W., 2021. Why you should stop predicting customer churn and start using uplift models. *Information Sciences*, 548, pp.497-515.

[19] Vo, N.N., Liu, S., Li, X. and Xu, G., 2021. Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, 212, p.106586.

[20] Pustokhina, I.V., Pustokhin, D.A., Aswathy, R.H., Jayasankar, T., Jeyalakshmi, C., Díaz, V.G. and Shankar, K., 2021. Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms. *Information Processing & Management*, 58(6), p.102706.

[21] Vijaya, J. and Sivasankar, E., 2019. An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Cluster Computing*, 22(5), pp.10757-10768.

[22] Jeyakarthic, M. and Venkatesh, S., 2020. An Effective Customer Churn Prediction Model using Adaptive Gain with Back Propagation Neural Network in Cloud Computing Environment. *Journal of Research on the Lepidoptera*, 51(1), pp.386-399.

[23] De Bock, K.W. and De Caigny, A., 2021. Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling. Decision Support Systems, p.113523.

[24] Li, Y., Hou, B., Wu, Y., Zhao, D., Xie, A. and Zou, P., 2021. Giant fight: Customer churn prediction in traditional broadcast industry. *Journal of Business Research*, 131, pp.630-639.

[25] Sivasankar, E. and Vijaya, J., 2019. Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network. *Neural Computing and Applications*, 31(11), pp.7181-7200.

[26] Vijaya, J. and Sivasankar, E., 2018. Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. *Computing*, 100(8), pp.839-860.

[27] Routh, P., Roy, A. and Meyer, J., 2021. Estimating customer churn under competing risks. *Journal of the Operational Research Society*, 72(5), pp.1138-1155.

[28] Pamina, J., Raja, J.B., Peter, S.S., Soundarya, S., Bama, S.S. and Sruthi, M.S., 2019, September. Inferring Machine Learning Based Parameter Estimation for Telecom Churn Prediction. In International *Conference On Computational Vision and Bio Inspired Computing (pp. 257-267). Springer, Cham.*

[29] Ahn, J., Hwang, J., Kim, D., Choi, H. and Kang, S., 2020. A Survey on Churn Analysis in Various Business Domains. *IEEE Access*, 8, pp.220816-220839.

[30] Gerber, G., Faou, Y.L., Lopez, O. and Trupin, M., 2020. The impact of churn on client value in health insurance, evaluation using a random forest under various censoring mechanisms. *Journal of the American Statistical Association*, pp.1-12.

[31] Amin, A., Al-Obeidat, F., Shah, B., Tae, M.A., Khan, C., Durrani, H.U.R. and Anwar, S., 2020. Just-in-time customer churn prediction in the telecommunication sector. *The Journal of Supercomputing*, 76(6), pp.3924-3948.

[32] Scriney, M., Nie, D. and Roantree, M., 2020, September. Predicting customer churn for insurance data. *In International Conference on Big Data Analytics and Knowledge Discovery (pp. 256-265). Springer, Cham.*

[33] Morik, K. and Köpcke, H., 2004, September. Analysing customer churn in insurance data–a case study. *In European conference on principles of data mining and knowledge discovery (pp. 325-336). Springer, Berlin, Heidelberg.*

[34] Rai, S., Khandelwal, N. and Boghey, R., 2020. Analysis of customer churn prediction in telecom sector using cart algorithm. *In First International Conference on Sustainable Technologies for Computational Intelligence (pp. 457-466). Springer, Singapore.*

[35] Mishachandar, B. and Kumar, K.A., 2018. Predicting customer churn using targeted proactive retention. *International Journal of Engineering & Technology*, 7(2.27), p.69.

[36] Jain, H., Khunteta, A. and Srivastava, S., 2020. Churn prediction in telecommunication using logistic regression and logit boost. *Procedia Computer Science*, 167, pp.101-112.

[37] Devriendt, F., Berrevoets, J. and Verbeke, W., 2021. Why you should stop predicting customer churn and start using uplift models. *Information Sciences*, 548, pp.497-515.

[38] Jin, H., 2022. The effect of overspending on tariff choices and customer churn: Evidence from mobile plan choices. *Journal of Retailing and Consumer Services*, 66, p.102914.

[39] Dalli, A., 2022. Impact of Hyperparameters on Deep Learning Model for Customer Churn Prediction in Telecommunication Sector. *Mathematical Problems in Engineering*, 2022.

[40] Shabankareh, M.J., Shabankareh, M.A., Nazarian, A., Ranjbaran, A. and Seyyedamiri, N., 2021. A Stacking-Based Data Mining Solution to Customer Churn Prediction. *Journal of Relationship Marketing, pp.1-24.*