International Journal of Intelligent Networks 4 (2023) 145–154

Contents lists available at ScienceDirect

International Journal of Intelligent Networks

journal homepage: www.keaipublishing.com/en/journals/international-journal-of-intelligent-networks

Customer churning analysis using machine learning algorithms

B. Prabadevi *, R. Shalini, B.R. Kavitha

School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

ARTICLE INFO

Keywords:
Stochastic gradient booster
Random forest
KNearest neighbors
Logistics regression
Customer churn
Machine learning

ABSTRACT

Businesses must compete fiercely to win over new consumers from suppliers. Since it directly affects a company's revenue, client retention is a hot topic for analysis, and early detection of client churn enables businesses to take proactive measures to keep customers. As a result, all firms could practice a variety of approaches to identify their clients early on through client retention initiatives. Consequently, this study aims to advise on the optimum machine-learning strategy for early client churn prediction. The data included in this investigation includes all customer data going back about nine months before the churn. The goal is to predict existing customers' responses to keep them. The study has tested algorithms like stochastic gradient booster, random forest, logistics regression, and k-nearest neighbors methods. The accuracy of the aforementioned algorithms are 83.9%, 82.6%, 82.9% and 78.1% respectively. We have acquired the most effective results by examining these algorithms and discussing the best among the four from different perspectives.

1. Introduction

The customer's concentration on the providers has prompted many new telecom associations to emerge. These new firms usually specialize in providing a specific service or product that the customer cannot find from the incumbent providers. These new firms can provide this service or product at a lower price than the incumbent providers, allowing them to capture a larger market share. The incumbent providers, however, can retain most of the market by pricing their products higher than the new firms. This competition between the incumbent providers and the new firms has caused the rates that the associations charge to change. The associations' rates are often determined by the amount of competition in the market. The more competition, the higher the rates the associations can charge. In markets with a low level of competition, the associations can charge rates lower than in markets with a high level of competition. The rates that the associations' charge is also affected by the number of services that the associations can offer. The more services the associations can offer, the higher the rates that the associations can charge.

Churning, in marketing terms, refers to the number of customers who stopped using a particular product. Always the churn rate must be low. Customer churning is common with any product when there are multiple options for a single problem. Usually, customers will churn when they face any difficulties or disappointments in the services rendered by the product. The churn rate is usually measured for a specific time. Any organization's primary motive should be satisfying customers and retaining existing customers. Retaining existing customers is equally important as gathering new customers. Customer churn prediction is the most important issue in adopting an industry's product.

Managing customer churn is one major challenge companies face, especially those offering subscription-based services. Customer churn also called customer attrition,n is the loss of customers, and it is caused by a change in taste, lack of proper customer relationship strategy, change of residence, and several other reasons. If businesses can effectively predict customer attrition, they can segment those customers that are highly likely to churn and provide better services to them. Hence, a churn prediction model is a mandate needed in today's digitized economy. An organization can achieve a high customer retention rate and maximize its revenue.

Among a few methodologies created in writing for anticipating client agitate, regulated Machine Learning (ML) procedures region the most and largely explored. ML consolidates many computations, for instance, Decision trees, KNearest Neighbors(KNN), Linear regression, Naive Bayes, Neural Networks, Support Vector Machines (SVM), Genetic Programming, and various others. Churn is one of those indispensable issues, and firms have begun to acquire new Business Intelligence (BI) applications that anticipate agitating clients. Whenever the association is familiar with the degree of clients who leave for one more organization in an extremely given specific period, it will be very simpler to return up with a nearby examination of the foundations for the agitate

* Corresponding author.
E-mail addresses: prabadevi.b@vit.ac.in (B. Prabadevi), shaliniravi1704@gmail.com (R. Shalini), kavitha.br@vit.ac.in (B.R. Kavitha).

https://doi.org/10.1016/j.ijin.2023.05.005
Received 8 September 2022; Received in revised form 17 April 2023; Accepted 21 May 2023
Available online 17 June 2023

rate and see the way of behaving of customers that withdraw and move to a various business competitor. This aids in planning compelling client maintenance strategies for that organization.

Such calculable existing estimations like Decision trees, KNN, Linear regression, Naive Bayes, Neural Networks, SVM, XG boost, and so forth are units accessible for client churning inside the market. Evidently, with these existing algorithms, it'll be troublesome to pick the most effective algorithms that suit our necessities. So in this study, we have analyzed some of these existing algorithms together on the same dataset. The best one suited for churning is determined based on the accuracy of these algorithms. Furthermore, these algorithms can improve the predictions' accuracy by incorporating the target variable's gradient. The study's main objective is to analyze the existing machine learning algorithms for earlier prediction of customer churn based on previously recorded information on customer feedback. Some potential churn prediction algorithms which identify the most important variables that affect the target variable are utilized in this study: Stochastic gradient booster, Random forest, K-Nearest Neighbors and Logistics Regression.

The rest of the sections in the paper is organized as follows: section 2 provides the survey on existing research work, section 3 details the proposed system, dataset description, and modules description, section 4 discusses the results and section 5 concludes the paper.

2. Literature survey

This section provides a briefly surveys search works related to customer churn prediction in various industries, their pros, and cons.

Omar Adwan et al., have used Multi-layer perceptron neural network (MLPNN): Modelling and Analysis tor to another commercial competitor, which leads to a loss of serious profits [1]. Actual customer data from a large Jordanian telecommunications company was provided for this investigation. Using MLP Neural Networks: Modelling and Analysis, they predict Customer Churn in the Telecommunications Industry.

Farhad Shaikh described the churn prediction system that uses classification and grouping techniques to rank churn clients and the reasons behind telecommunication customer churn [2]. They tend to project churn identification and prediction from an oversized telecommunications dataset, mistreatment of using ML and NLP techniques.

Babu and Ananth have proposed that it is the procedure of finding the information concealed in huge informational collections, incorporates different strategies and calculations to play out an effective investigation of informational indexes, the order is the strategy that is utilized to recognize the information and make forecasts about the future remain constant [3].

Ismail et al., have proposed that the media communications industry faces critical rivalry between merchants to draw in new client to your provider [4]. The study recommends a MLPNN way to deal with foreseen client agitate at one of Malaysia's driving media communications organizations. The outcomes are analyzed against the most well-known misfortune expectation procedures, like regression and classification. The experiments show that the neural network-based approach medium-sized NNs were found to perform higher in predicting client churn once experiencing different neural network topologies [5].

Kosgey have used text summarization for churn prediction techniques to realize a deeper perception of client churn and shows that the foremost correct churn prediction is provided hybrid victimization models instead of individual algorithms to assist the telecommunications trade in understanding client churn wants and improve their services undo the choice to cancel [6].

Fatih Kayaalp, has proposed that the churn analysis is one of all the analyses used worldwide in subscription-oriented industries to research client behavior and predict customers UN agency can leave the mass service of an organization [7]. To keep the review up thus far, studies revealed within the past five years, primarily within the past two years, are enclosed.

Gholamiangonabadi et al. proposed that this article introduces a brand-new approach to measure the churn rate at Iranian banks: initially, it is normalized by data preparation [8]; then, the information pool is formed employing a k-medoids method. The results recommend that the MLPNN and SVM models have higher preciseness and lower value practicality.

Amuda and Adeyem have described a prognosticate model that uses the multilayer perceptron design of the Artificial Neural Network (ANN) to predict client churn in an exceeding institution. The results showed that the event of the ANN code had a performance admire of the Neuro answer eternity code [9].

Amit had surveyed many shoppers' information and states that customer churninig becomes tougher as they can't target every customer's needs regarding services. Once customers needs are not fulfilled, they switch from the service supplier [10]. Anam has proposed that customer Churn has one of the foremost necessary telecommunications issues [11]. This text provides an outline of varied data processing techniques for churn prediction.

Deepthi Das and Raju Ramakrishna have proposed that numerous business and e-commerce specialists know customer Churn to spot customers in United Nations agency area units ready to amend their existing business service or finish their subscription term [12]. Recently, corporations like e-commerce, telecommunications, and insurance have returned below tremendous pressure.

ULLAH et al. have developed a model whose outcomes show that the proposed stir expectation model gave better agitate order utilizing the RF calculation and the client profile using k-means grouping [13]. In expansion, it also gives beat variables to the standards produced using the chosen characteristic classifier calculation.

The best prescient model accomplished 79% of the mathematical mean, and misclassification mistakes were limited to 0.192. 0.229 for Type I or Type II blunders. In outline, a fascinating Meta Cost strategy has worked on presenting the visionary model without requiring critical handling by changing the first information tests [14].

Edvaldo and Olawande have proposed that as of not long ago, conventional AI strategies (for example, MLP and SVM were effectively utilized for pivot forecast, however with impressive exertion in the design of the preparation boundaries [15]. They were foreseeing even information for client relationships with the executives in finance utilizing Deep Neural Networks (DNN).

Yahaya and Abisoye have proposed that the client stir expectation is a significant issue in the financial industry and has acquired consideration over the years [16]. The execution shows that the preparation execution further developed when a commotion was sifted, while the testing execution was impacted by the unequal information brought about by separating.

Ahmed and Shabib Aftab suggested an approach that require some investment and assets to create top-notch programming because main modules that are anticipated to be imperfect are tried [17]. They have presented an order system that utilizes the Multi Filter highlight choice procedure, and MLP is used to foresee blunder-inclined programming modules. As indicated by the outcomes, the structure proposed utilizing the class harmony method functioned admirably in all the datasets.

Amatare and Adebola OJO, described that this study proposed a Convolution neural network (CNN) model for anticipating client beat in a media communications industry; three different models fostered the utilization of two MLP models and another CNN model [18]. Precision rates for MLP models; MLP1 and MLP2 are 80% and 81%, respectively, while the CNN, CNN1, and CNN2 models are 81% and 89%, respectively.

M.Feindta et al., have proposed that the detailed analysis related to knowledge plays a vital role in fashionable analytics. Mechanically preprocesses input variables and uses advanced regularization and clipping techniques to el primarily eliminate the danger of overtraining [19]. Sun-Chong et al., considers artificial neural networks as approximations of universal functions. We tend to gift the neural networks that area unit appropriate for statistic forecasts [20].

Edwine et al. [21], have done a comparative analysis of customer churn prediction models in the telecom industry. They have used three best-fit algorithms, namely KNN, RF, and SVM, along with an optimization algorithm for hyperparameter tuning. They have concluded that the basic versions of these algorithms perform lesser than the amalgamation (RF with grid search optimization algorithm) with a low-ratio undersampling strategy. Similarly, a few more works on improving customer churn prediction are performed in telecom customer segmentation using logistic regression [22] and suggestions on optimized solutions to the existing machine learning models with a focus on feature reduction (an optimized subset of features to predict the model) [23]. Furthermore, other prediction models that can be suggested for any generic predictive analytic applications are discussed in Refs. [24,25].

It has been evident from the survey that machine learning and artificial intelligence play a wider role in customer churn analysis. Furthermore, it has been observed that machine learning algorithms perform better in integration than individual performance. Deep learning algorithms are preferred only for image-based reviews and sentiment analysis. Also, optimization of the model through feature engineering is suggested. Therefore, an analysis of the best-fit algorithms for customer churn prediction using machine learning is performed in this paper to assist readers and researchers. The results of this study provide useful insights into the industry and help them predict customer churn at an early stage and retain customers. Logistic regression better interprets the data. K-Nearest Neighbors offers accurate predictions. Random Forest is preferred as it splits the trees based on a subset of features to improve the bagging. Furthermore, the Stochastic Gradient Booster is preferred as it checks the training samples at random instead of all and helps to reach the global minimum faster.

3. Proposed analysis

This section describes the details of the study, methodologies used, and modules.

3.1. Study flow

The concept of improvement is finding the most effective answer for future problems by gaining expertise from the present examples within the machine learning method. Customer churn prediction has been performed using different methods, techniques, data processing, machine learning, and hybrid technologies. Most of them used call trees because it is a recognized way to seek out client churn. However, it does not apply to complicated issues. However, the study shows that reducing the information improves the accuracy of the decision tree. In some cases, machine learning algorithms square measure used for client prediction and historical analysis.

Here projected review methodology consists of a few phases, as depicted in Fig. 1. The dataset obtained from the Kaggle with 7044 alternatives and 21 attributes was taken as input. Within the 1st 2 phases, knowledge preprocessing and analysis are performed. Then, the information has been parted into two sections, train,and test set in the proportion of 70% and 30% separately. The most well-liked prophetical models are applied within the prediction method: Logis Regression, KNN, random gradient booster, random forest, etc., and ensemble techniques square measure applied to visualize the impact on the accuracy of models. Finally, a giant knowledge platform should create the churn prophetical system curve.

- To create the churn prophetical system, a giant knowledge platform should be put in. Jupyter libraries were picked because it is a free and open-source structure.
- The importance of this kind of analysis within the market is to assist organizations in building much profit.
- Overall, the findings recommend that a neural network learning algorithmic program may provide a viable various to applied math prophetical approaches in client churn prediction.
- The final results of all the algorithmic program can show that algorithmic program is best appropriate for churn prediction

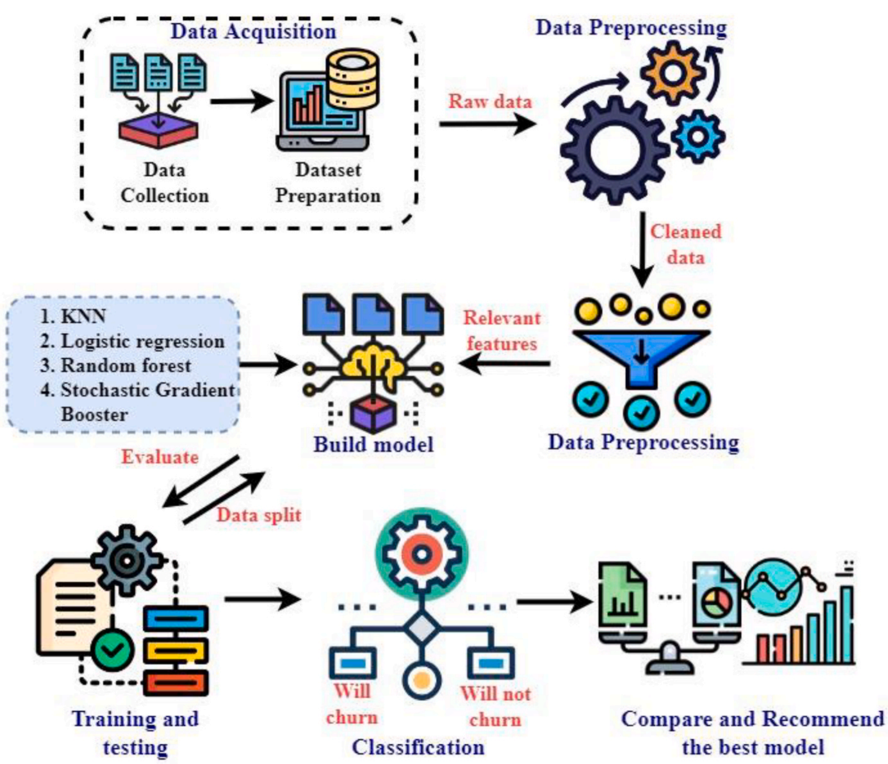

Fig. 1. System layout.

3.2. Methodologies used

The system involved in the analysis of customer churning uses four different algorithms mentioned below.

1. Stochastic gradient booster
2. Random forest model
3. K-Nearest Neighbors
4. Logistic regression model

3.2.1. Stochastic gradient Booster(SGB)

This variation of boosting is termed random gradient boosting. SGB architecture is depicted in Fig. 2. At every iteration, a subsample of the training data is chosen randomly (without substitution) from the full preparation dataset. The arbitrarily chosen subsample is then utilized to fit the base student rather than the full example. A couple of stochastic variations can be utilized: Subsample columns before making each tree. Subsample segments before making each tree.

Steps concerned in SGB as specified in [17].

Training rule.

1. Resulting to start the training instructing
2. Loads
3. Inclination
4. For simple estimation and straightforwardness, NN parameters like biases and weights ought to be set to satisfactory zero, and the learning rate ought to be set to sufficient.
5. Activate each info unit as follows $x_i = s_i(i = 1\text{to } n)$
6. Get web input with the resulting ensuing.
7. Apply appropriate activation function to determine a definitive from the results in step 6.

Purpose of stochastic gradient booster.

- Gradient boosting algorithmic may be used to predict nonstop objective variables (as a Regressor) but additionally categorical target variables (as a Classifier).
- As gradient boosting is one of the boosting algorithms, it's accustomed to minimizing the bias error of the model.
- When it's used as a regressor, the price performed is Mean square Error (MSE), and once it's used as a classifier, then the price performed is Log loss

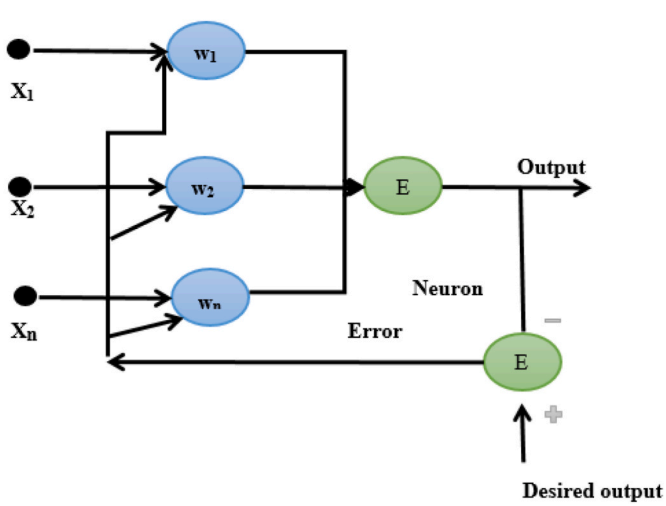

Fig. 2. Stochastic Gradient Booster Architecture [17].

- Loads of adaptability - can improve on various misfortune works and gives a few hyper boundary tuning choices that make the capacity fit entirely adaptable.
- No information preprocessing is needed - typically works nicely with categorical and numerical values as is.

Challenges.

- One detriment of supporting is that it is delicate to exceptions since each classifier is obliged to fix the mistakes in the ancestors.
- Along these lines, the technique is too reliant upon exceptions. Another weakness is that the technique is difficult to increase.
- This is because each assessor puts together its rightness concerning the past indicators, accordingly making the methodology hard to smooth out.
- The high adaptability brings about numerous boundaries that cooperate and impact intensely the way of behaving of the methodology (number of cycles, tree profundity, regularization boundaries, and so forth) This requires an enormous framework search during tuning.
- Less informative, though this can be simply self-addressed with numerous tools.

3.2.2. Random forest model (RF)

Random Forest is appropriate for a large dataset. Random Forest architecture is depicted in Fig. 3. A technique that mixes several classifiers to supply solutions to complicated issues. Similarly, a random forest algorithmic program creates call trees on information sum, gets the prediction from every one of them, and eventually selects the simplest answer by suggesting that of the ballot.

Steps concerned in random forest algorithm.

1. n scope of arbitrary records square measure taken from the data set having k scope of records.
2. Individual decision trees are worked for every model.
3. Each decision tree will make an outcome.

The outcome is separately established on Majority Voting or Averaging for Classification and backsliding.

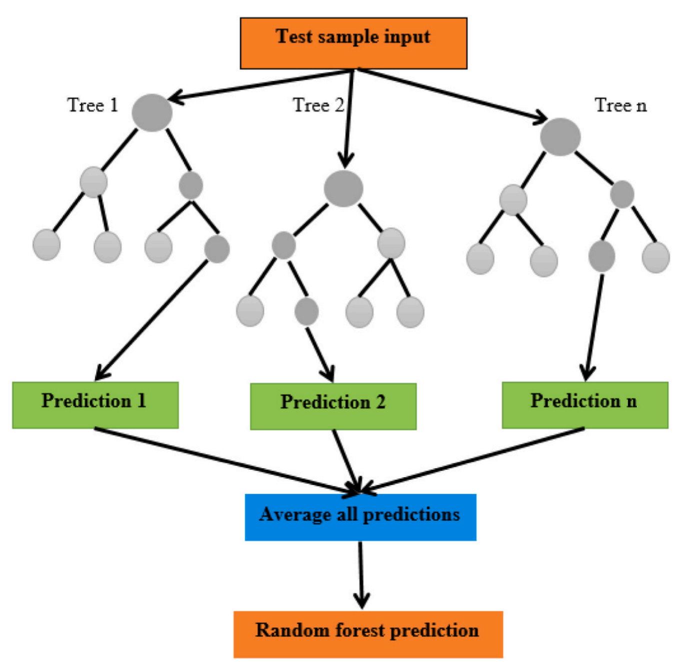

Fig. 3. Random Forest Architecture [13].

Challenges.

- **Time-consuming process:** Since random forest algorithms will handle massive information sets, they will be offered additional correct predictions However, they will be slow to method information as they're computing information for every individual call tree.
- **Requires additional resources:** Since the random forests method has larger information sets, they'll need other help to store that information.
- **More complex:** The prediction of one call tree is simpler to interpret than a forest of them.

Purpose of random forest.

- Among all the accessible arrangement techniques, random forests provide the most remarkable precision.
- The random forest technique can even handle huge information with varied variables running into thousands.
- It will mechanically balance information sets once a category is a lot more than different categories within the information.
- At the center of this calculation is a Decision Tree along these lines; Random Forests share every one of its benefits.

3.2.3. KNN

It's a supervised machine learning rule. KNN architecture is depicted in Fig. 4. The calculation can be utilized to take care of both regression and classification issue articulations. KNN calculation at the preparation stage merely stores the dataset. Once it gets new info, it arranges that info into a classification.

KNN algorithmic rule [4]

1. Load the information
2. Initialize K For every model in the information
3. Sort the arranged assortment of distances and files from smallest to largest (in ascending order) by the distances
4. Get the labels of k
5. If Regression, come back the mean of the K labels
6. If classification, come back the mode of the K labels

Challenges.

- **Easy to implement:** Given the algorithm's simplicity and accuracy, it's one of the primary classifiers that a brand-new information human can learn.

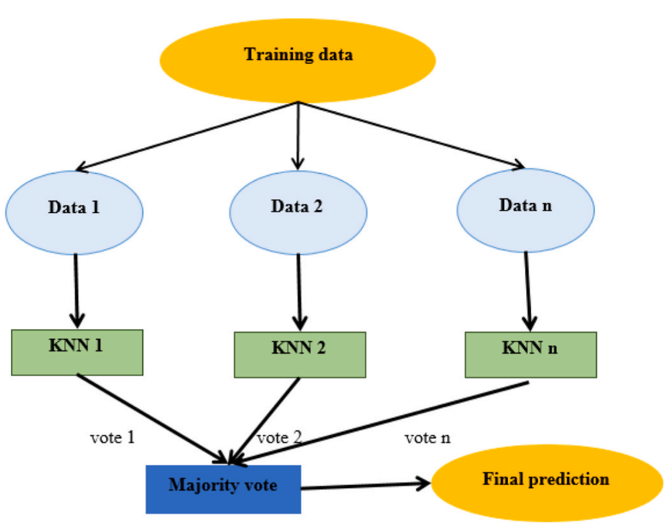

Fig. 4. KNN architecture [4].

- **Adapts easily:** As new coaching samples square measure extra, the algorithmic rule adjusts to account for any new information since all coaching information is held on into memory.
- **Few Hyperparameters:** KNN just requires a k-worth and a distance metric, which is low when contrasted with other AI calculations.
- **High accuracy** –you have to contrast and worse regulated learning models
- **No assumptions concerning information** need to be compelled to create further assumptions, tune many parameters, or build a model. This makes it crucial in nonlinear information cases.

Purpose of the KNN algorithm.

- KNN is utilized in each regression and classification of prophetical issue. It uses information with many categories to predict the classification of the new sample purpose.
- KNN is exceptionally simple to execute as the main thing to be determined is the distance between various focuses based on information of various elements and this distance can without much of a stretch be determined utilizing distance recipes, for example, Euclidian or Manhattan.
- As there is no preparation period, new information can be added whenever since it won't influence the model.
- Accuracy depends on the standard of the information. With massive data, the prediction stage may well be slow.
- Sensitive to the dimensions of the information and tangential options.
- Require high memory – need to store all of the preparation information.

3.2.4. Logistic regression model

The logistic Regression can classify our observations because the client "will churn" or "won't churn" from the platform. This architecture is depicted in Fig. 5. This model will attempt to figure out the likelihood of happiness in at least one cluster or another.

Steps carried in LR calculation.

Step 1. Import the necessary libraries
Step 2. Peruse and get the information
Step 3. Exploratory Data Analysis
Step 4. Information Preparation
Step 5. Building Logistic Regression Model
Step 6. Making Predictions on Test Set
Step 7. Appointing Scores according to anticipated likelihood values

Challenges.

- The major challenge of logistical Regression is the assumption of dimensionality between the variable quantity and the freelance variables.
- Nonlinear issues cannot be solved with logistical Regression since it's a linear call surface.

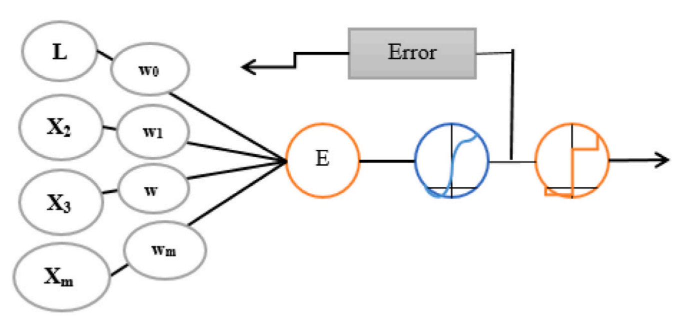

Fig. 5. Logistic Regression Architecture [22].

- Therefore, the transformation of nonlinear options is needed, which may be done by increasing the number of elements such that the information becomes linearly dissociable in higher dimensions.

Purpose of Logistic Regression.

- Logistic Regression predicts the output of a categorical variable quantity. So the result should be a categorical or distinct worth. It will be either affirmative or no, 0 or 1, true or False, etc. however, rather than giving the precise worth as zero and one, it provides the probabilistic values that lie between zero and one.
- It will simply reach multiple classes (multinomial Regression) and a natural probabilistic perspective on class expectations
- Logistic Regression can be utilized to group the perceptions utilizing various sorts of information and can,without much of a stretch, decide the best factors utilized for the arrangement.
- Logistic Regression performs higher once the information is linearly dissociable
- It doesn't need too several procedure resources as its extremely explainable. There is no downside to scaling the input features—It doesn't need standardization
- It is simple to implement and train a model victimization logistical regression
- It provides a live-off, however relevant a predictor (coefficient size) is, and its direction of association (positive or negative).

Dataset Description.

The customer churn dataset was downloaded from Kaggle. The customer information contains data about a made-up telco organization that provided home telephone and Internet administrations to 7044 clients. It demonstrates which clients have left, remained, or pursued their administration. While gazing at knowledge from customers that have already got churned (response) and their characteristics/behavior (predictors) before the churn happened, these information were recorded.

The informational index incorporates data about.

1 Clients who left inside the last month - the portion is called Churn
2 Administrations that each client has pursued - telephone, numerous lines, web, online security, online reinforcement, gadget assurance, school backing, and streaming TV and flicks
3 client account information - how long they've been a client, contract, installment technique, paperless charging, month-to-month charges, and outright charges
4 Segment data concerning clients - orientation, age range, and on the off chance that they need accomplices and wards.

The attributes of the dataset with 7044 alternatives are given in Table 1.

4. Result and discussion

The results were obtained using Python 3.10.6 by utilizing the Jupyter Libraries from Anaconda. The various libraries used include numpy, pandas, matplotlib and seaborn.The results obtained in comparing the performance of the various algorithms are narrated step by step.

4.1. Test and train dataset split

The customer churn dataset is split into training and testing data in a 70 : 30 ratio respectively. The head of the dataset is shown in in Fig. 6.

Table 1
Dataset attributes and description.

S. No	Attribute Name	Description
1.	Customer id	The id of the customer
2.	Gender	The customer's or client's gender, whether male or female
3.	Senior-resident	Whether the customer or client is a senior resident or not (1, 0)
4.	Partner	Whether the client or the customer has a collaborator or not (Yes, No)
5.	Dependents	Whether the client or the customer has wards/family members as dependants or not (Yes, No)
6.	Tenure	Number of months the client or the customer has retained with the organization
7.	Phone Service	Whether the client or the customer has a telephone/ mobile connection or not (Yes, No)
8.	Multiple Lines	Whether the customer or client has several telephone lines or not (Yes, No, No telephone administration)
9.	Internet Service	Customer's web access supplier (DSL, Fiber optic, No)
10.	Online Security	Whether the customer or client has web security or not (Yes, No, No web access)
11.	Online Backup	Whether the customer or client has online reinforcement or not (Yes, No, No network access
12.	Device Protection	Whether the customer or the client has gadget assurance or not (Yes, No, No web access)
13.	Tech Support	Whether the customer or the client has technical support or not (Yes, No, No network access)
14.	Streaming TV	Whether the customer or the client has streaming TV or not (Yes, No, No network access)
15.	Streaming Movies	Whether the customer or the client has streaming motion pictures or not (Yes, No, No web access)
16.	Contract	The customer's or the client's agreement term (Month-to-month, One-year, long-term)
17.	Paperless Billing	Whether the customer or the client has paperless charging or not (Yes, No)
18.	Payment Method	The customer's or the client's installment strategy
19.	Total Charges	The aggregate sum
20.	Monthly Charges	The aggregate charged to the client every month
21.	Agitate	If the customer or the client stirred or not (Yes or No)

4.2. Dataset and its description before and after datatype conversion

The raw dataset obtained has the attributes in different types of data like objects. So the data is categorized and converted to a feasible type. The dataset description for training with different models is displayed in Fig. 7a. and Fig. 7b. One hot encoding and label encoders were used to transform the categorical labels to numerical labels and for normalizing the labels as shown in Fig. 7c.

4.3. Prediction of the KNN algorithm

The KNN cross-validation is performed using Grid Search CV for the hyperparameter tuning, and the best KNN results after tuning are specified. The prediction results of KNN algorithm are depicted in Fig. 8a and b. The best KNN Training score was 0.7849459340352903, the test performance was 0.7773049645390071 and the AUROC was 0.7817042078747963.

4.4. Prediction of logistic regression algorithm

The prediction results of the LR algorithm by applying GridSearch CV for hyperparameter tuning are displayed in Fig. 9a. and 9b. The best LR Training Score was 0.7975697081209744, test Performance was 0.7829787234042553 and AUROC was 0.8269877975760328.

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

Fig. 6. The sample rows of the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
None
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   int64
 1   gender            7043 non-null   int64
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   int64
 4   Dependents        7043 non-null   int64
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   int64
 7   MultipleLines     7043 non-null   int64
 8   InternetService   7043 non-null   int64
 9   OnlineSecurity    7043 non-null   int64
 10  OnlineBackup      7043 non-null   int64
 11  DeviceProtection  7043 non-null   int64
 12  TechSupport       7043 non-null   int64
 13  StreamingTV       7043 non-null   int64
 14  StreamingMovies   7043 non-null   int64
 15  Contract          7043 non-null   int64
 16  PaperlessBilling  7043 non-null   int64
 17  PaymentMethod     7043 non-null   int64
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   float64
 20  Churn             7043 non-null   int64
dtypes: float64(2), int64(19)
memory usage: 1.1 MB
None
```

Fig. 7a. Dataset description.

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590	1	0	1	0	1	0	3	1	0	...
1	5575	2	0	0	0	34	1	0	1	1	...
2	3668	2	0	0	0	2	1	0	1	1	...
3	7795	2	0	0	0	45	0	3	1	1	...
4	9237	1	0	0	0	2	1	0	2	0	...

5 rows × 21 columns

Fig. 7b. Dataset samples after conversion.

4.5. Prediction of random forest algorithm

The prediction results of the RF algorithm by applying RandomizedSearch CV for hyperparameter tuning are displayed in Fig. 10a. and 10b. The best RF Training Score was 0.8038805992445953, test Performance was 0.7872340425531915 and AUROC was 0.829097309685545.

4.6. Prediction of stochastic gradient booster algorithm

The prediction results of the SGB algorithm by applying RandomizedSearch CV for hyperparameter tuning are displayed in Fig. 11a. and

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines
customerID	1.000000	0.006073	-0.002197	-0.026779	-0.012816	0.007805	-0.006252	0.007550
gender	0.006073	1.000000	-0.001874	-0.001808	0.010517	0.005106	-0.006488	0.001806
SeniorCitizen	-0.002197	-0.001874	1.000000	0.016479	-0.211185	0.016567	0.008576	0.071049
Partner	-0.026779	-0.001808	0.016479	1.000000	0.452676	0.379697	0.017706	0.061417
Dependents	-0.012816	0.010517	-0.211185	0.452676	1.000000	0.159712	-0.001762	-0.011900
tenure	0.007805	0.005106	0.016567	0.379697	0.159712	1.000000	0.008448	0.176459
PhoneService	-0.006252	-0.006488	0.008576	0.017706	-0.001762	0.008448	1.000000	-0.844955
MultipleLines	0.007550	0.001806	0.071049	0.061417	-0.011900	0.176459	-0.844955	1.000000
InternetService	-0.012230	-0.000863	-0.032310	0.000891	0.044590	-0.030359	0.387436	-0.381534
OnlineSecurity	-0.000409	-0.000214	-0.208709	0.056157	0.179614	0.085500	0.146522	-0.249780
OnlineBackup	-0.006665	0.000788	-0.170002	0.059540	0.161106	0.107643	0.164540	-0.244690
DeviceProtection	-0.008100	0.005642	-0.172926	0.064584	0.157003	0.107656	0.156631	-0.237032
TechSupport	-0.004863	0.002805	-0.217566	0.047420	0.173036	0.084902	0.145215	-0.247977
StreamingTV	-0.008606	0.002992	-0.155266	0.054605	0.146505	0.078087	0.179510	-0.246600
StreamingMovies	-0.012090	0.002082	-0.149000	0.051632	0.136652	0.081169	0.175257	-0.241952

Fig. 7c. Sample Dataset after Label encoding and One hot encoding.

```
KNN AUROC: 0.7814042078747963
              precision    recall  f1-score   support

           0       0.81      0.92      0.86       518
           1       0.63      0.39      0.48       187

    accuracy                           0.78       705
   macro avg       0.72      0.65      0.67       705
weighted avg       0.76      0.78      0.76       705
```

Fig. 8a. Final accuracy of KNN.

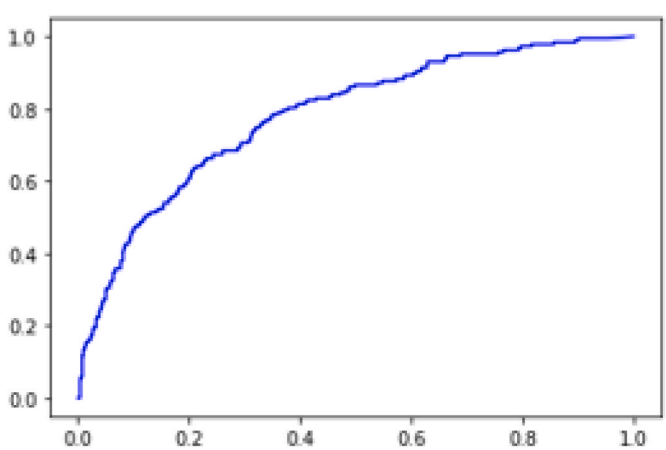

Fig. 8b. Predicted graph for KNN.

```
Tuned LR Parameters: {'C': 0.05}
Best LR Training Score:0.8002514696033005
 LR Test Performance: 0.7843971631205674
 LR AUROC: 0.8176036999566413
```

Fig. 9a. Final accuracy of LR.

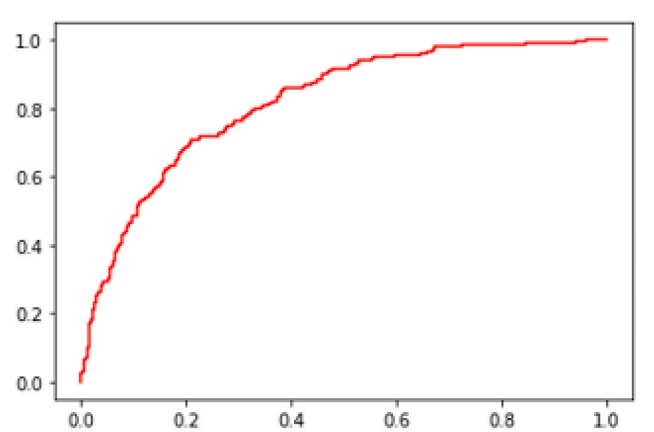

Fig. 9b. Predicted graph for LR.

```
RF AUROC: 0.839097309685545
              precision    recall  f1-score   support

           0       0.82      0.91      0.86       518
           1       0.64      0.46      0.53       187

    accuracy                           0.79       705
   macro avg       0.73      0.68      0.70       705
weighted avg       0.77      0.79      0.78       705
```

Fig. 10a. Final accuracy of Random Forest.

11b. The best SGB Training Score was 0.8067218322921829, test Performance was 0.7914893617021277 and AUROC was 0.8396754279107219.

4.7. Prediction results of overall compared four algorithms

The four algorithms taken for the analysis were described in Table 2.

The four algorithms under study were compared based on the accuracy as shown in Fig. 12. It is shown that SGB performs better than other models. The performance metrics comparison of all four models is

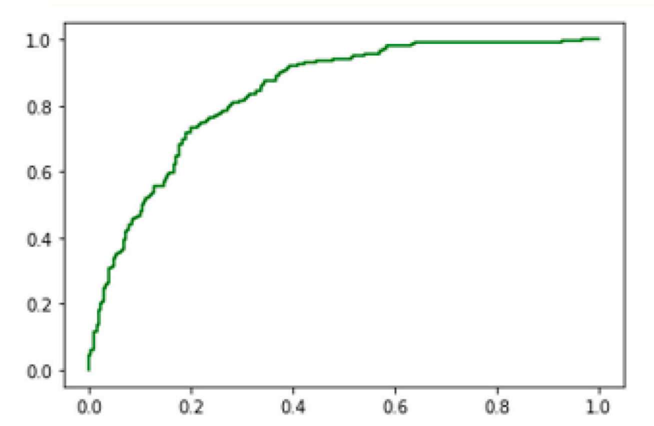

Fig. 10b. Predicted graph for Random Forest.

```
SGB AUROC: 0.8396754279107219
              precision    recall  f1-score   support

           0       0.82      0.91      0.87       518
           1       0.65      0.46      0.54       187

    accuracy                           0.79       705
   macro avg       0.74      0.69      0.70       705
weighted avg       0.78      0.79      0.78       705
```

Fig. 11a. Final accuracy of stochastic gradient booster.

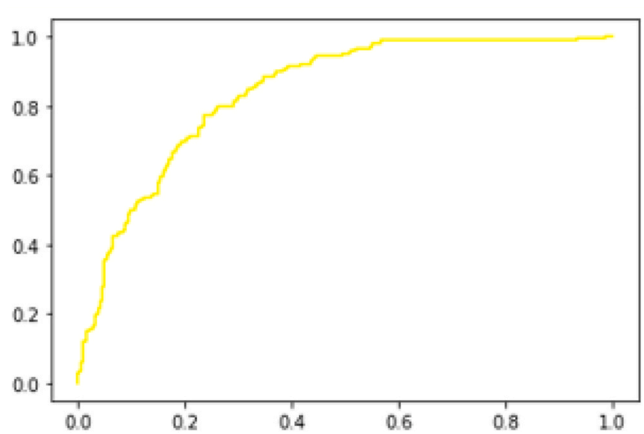

Fig. 11b. Predicted Graph for stochastic gradient booster.

shown in Fig. 13.

The models chosen are examined through ROC and AUC. ROC show the yes instead of the Misleading Positive Rate of matched classifiers across every decision edge (someplace in the scope of 0, 1). In the upper right is any place the decision edge is zero. In this way, any perception with a P (y = 1)≥zero is classed as a "1″, and the rest region unit is delegated a "0". Since each perception can fulfill the essential condition, every perception is classed as a "1". Subsequently, we will quite often appropriately characterize all obvious "1″s, but inaccurately order all evident "0″s.

Therefore, organizations can use SGB to predict customer churning rates better, increasing retention rates and reducing costs incurred. This model can be integrated into the organization's customer management portal to monitor the customer and predict the retention rate. This will help categorize the customers as at-risk and can further improve to satisfy their needs.

Customer churning for a different context and on different datasets was carried out by considering other factors like Social network analysis

Table 2
The difference among Logistic Regression, KNN, Stochastic gradient booster and Random Forest.

Factor	Logistic regression	K-Nearest Neighbors	Stochastic gradient booster	Random forest
Definition	Used to predict the categorical dependent value and solve a classification problem	It consumes more cost for training. It is non-parametric model	Uses delta rule for training Weights and bias are adjustable	Used in classifications and regressions problems Suitable for large data and interpretability is not a major concern
Accuracy	**0.826**	**0.781**	**0.839**	**0.829**
Advantage	Easier to implement and interpret and efficient to train	Much faster than other algorithms for training	Its efficiency is one of the major advantages	provide the highest accuracy, but in this work SGB outperforms
Remark	Better than others	Gives decent value	Gives the best performance compared with the other three	Second best performance

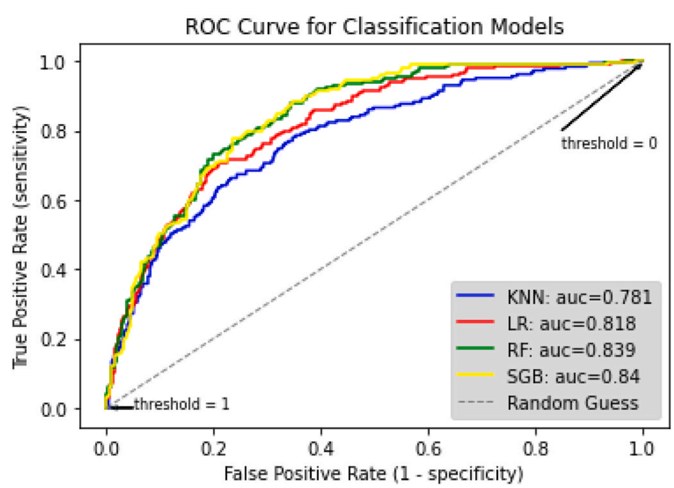

Fig. 12. ROC Curve for classification Models.

Fig. 13. Comparative analysis of KNN, LR, RF, and SGB.

features. The results obtained on the unbalanced dataset category for RF are 6% less than the proposed model [26]. Furthermore, better accuracy was obtained using a backpropagation network and decision tree, but this combination of machine learning algorithms was least tried [27]. Lalwani et al. have developed a prediction model by applying machine learning algorithms like LR, Naïve Bayes, support vector machine, RF, DT, ensemble methods, and cross-validation for hyperparameter tuning. The highest accuracy of 81.71% and 80.8% were obtained by the AdaBoost and XGBoost classifiers, respectively [28], which is lesser than the highest accuracies of the proposed model.

5. Conclusions and future works

The results were examined to observe the exhibition regarding the different calculations of planning data for customer churn analysis. It had become understood that foreseeing a stir is one in everything about preeminent essential wellsprings of monetary benefit to partnerships. Four algorithms were chosen due to their variety during this prediction and are examined by investigating Receiver Operating Characteristics (ROC) and Area Under Curve(AUC). Looking at the regions beneath the bend, the Stochastic Gradient Booster model plays out magnificent, with an AUC of 0.84; the most over-the-top dreadful model is the K-Nearest Neighbor, which is a still-fair AUC of 0781. The Grid Search CV has increased the time needed to figure hyperparameter optimization. Furthermore, looking over a bigger hyperparameter space doesn't ensure that Randomized Search CV will choose the ideal hyperparameters.

Further research can concentrate on refined data-side preprocessing and exhaustive hyperparameter standardization to improve the model performance. More advanced optimization methods can be used for hyperparameter optimization. To achieve the most accuracy despite machine intensity, a better hyperparameter optimization method would probably increase the classification accuracy of the models to some extent.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Omar Adwan, Hossam Faris, Khalid Jaradat, Osama Harfoushi, Nazeeh Ghatasheh, Predicting customer churn in telecom industry using multilayer preceptron neural networks: modeling and analysis, Life Sci. J. 11 (3) (2014) 75–81.
[2] Mohammad Ridwan Ismail, Mohd Khalid Awang, M. Nordin A. Rahman, Mokhairi Makhtar, A multi-layer perceptron approach for customer churn prediction, International Journal of Multimedia and Ubiquitous Engineering 10 (7) (2015) 213–222.
[3] Farhad Shaikh, Brinda Pardeshi, Ajay Jachak, Akash Bendale, Nandakishor Sonune, Mangal Katkar, Customer churn prediction using nlp and machine learning: an overview, International Journal Of Advance Scientific Research And Engineering Trends 6 (2) (2021) 40–45.
[4] Anuj Sharma, DrPrabin Kumar Panigrahi, A neural network based approach for predicting customer churn in cellular network services, Int. J. Comput. Appl. 27 (11) (2011) 26–31.
[5] Ammara Ahmed, D. Maheswari Linen, A review and analysis of churn prediction methods for customer retention in telecom industries, in: 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, 2017, pp. 1–7.
[6] S. Babu, N.R. Ananthanarayanan, V. Ramesh, A study on efficiency of decision tree and multi layer perceptron to predict the customer churn in telecommunication using WEKA, Int. J. Comput. Appl. 140 (4) (2016) 26–30.
[7] Fatih Kayaalp, Review of customer churn analysis studies in telecommunications industry, Karaelmas Science & Engineering Journal 7 (2) (2017).
[8] Davoud Gholamiangonabadi, Jamal Shahrabi, Seyed Mohamad Hosseinioun, Sanaz Nakhodchi, Soma Gholamveisy, Customer churn prediction using a new criterion and data mining; A case study of Iranian banking industry, in: Proceedings of the International Conference on Industrial Engineering and Operations Management, 2019, pp. 5–7.
[9] Kamorudeen A. Amuda, Adesesan B. Adeyemo, Customers Churn Prediction in Financial Institution Using Artificial Neural Network, 2019, 11346 arXiv preprint arXiv:1912.
[10] Anam Bansal, Churn prediction techniques in telecom industry for customer retention: a survey, J. Eng. Sci. 11 (4) (2020) 871–881.
[11] Vrushabh Jinde, Savyanavar Amit, " customer churn prediction system using machine learning.", International Journal of Advanced Science and Technology 29 (5) (2020) 7957–7964.
[12] P. Sri Sai Surya, K. Anitha, Comparative analysis of accuracy and prediction of customer loyalty in the telecom industry using novel diverse algorithm, in: 2022 International Conference on Business Analytics for Technology and Security (ICBATS), IEEE, 2022, pp. 1–7.
[13] Irfan Ullah, Basit Raza, Ahmad Kamran Malik, Muhammad Imran, Saif Ul Islam, Sung Won Kim, A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector, IEEE Access 7 (2019) 60134–60149.
[14] Nazeeh Ghatasheh, Hossam Faris, AlTaharwa Ismail, Yousra Harb, Ayman Harb, Business analytics in telemarketing: cost-sensitive analysis of bank campaigns using artificial neural networks, Appl. Sci. 10 (7) (2020) 2581.
[15] Edvaldo Domingos, Blessing Ojeme, Olawande Daramola, Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector, Computation 9 (3) (2021) 34.
[16] Rahmat Yahaya, Opeyemi Aderiike Abisoye, Sulaimon Adebayo Bashir, An enhanced bank customers churn prediction model using A hybrid genetic algorithm and K-means filter and artificial neural network, in: 2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA), IEEE, 2021, pp. 52–58.
[17] Ahmed Iqbal, Shabib Aftab, A classification framework for software defect prediction using multi-filter feature selection technique and MLP, Int. J. Mod. Educ. Comput. Sci. 12 (1) (2020).
[18] Sunday A. Amatare, A.K. Ojo, Predicting customer churn in telecommunication industry using convolutional neural network model, IOSR J. Comput. Eng. 22 (3) (2020) 54–59.
[19] M. Feindt, U. Kerzel, The NeuroBayes neural network package, Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip. 559 (1) (2006) 190–194.
[20] Sun-Chong Wang, Artificial neural network, in: Interdisciplinary Computing in Java Programming, Springer, Boston, MA, 2003, pp. 81–100.
[21] Nabahirwa Edwine, Wenjuan Wang, Wei Song, Denis Ssebuggwawo, Detecting the risk of customer churn in telecom sector: a comparative study, Math. Probl Eng. 2022 (2022). Article ID 8534739, 16 pages.
[22] Tianyuan Zhang, Sérgio Moro, Ricardo F. Ramos, A data-driven approach to improve customer churn prediction based on telecom customer segmentation, Future Internet 14 (3) (2022) 94.
[23] Seyed Mohammad Sina Mirabdolbaghi, Babak Amiri, Model optimization analysis of customer churn prediction using machine learning algorithms with focus on feature reductions, Discrete Dynam Nat. Soc. 2022 (2022). Article ID 5134356, 20 pages.
[24] Narges Heidari, Parham Moradi, Abbas Koochari, An attention-based deep learning method for solving the cold-start and sparsity issues of recommender systems, Knowl. Base Syst. 256 (2022), 109835, https://doi.org/10.1016/j.knosys.2022.109835. ISSN 0950-7051.
[25] Navid Khaledian, Farhad Mardukhi, CFMT: a collaborative filtering approach based on the nonnegative matrix factorization technique and trust relationships, J. Ambient Intell. Hum. Comput. (2022) 1–17.
[26] A.K. Ahmad, A. Jafar, K. Aljoumaa, Customer churn prediction in telecom using machine learning in big data platform, Journal of Big Data 6 (1) (2019) 1–24.
[27] Thanasis Vafeiadis, Konstantinos I. Diamantaras, Sarigiannidis George, K. Ch Chatzisavvas, A comparison of machine learning techniques for customer churn prediction, Simulat. Model. Pract. Theor. 55 (2015) 1–9.
[28] Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, Pratyush Sethi, Customer churn prediction system: a machine learning approach, Computing (2022) 1–24.