Performance analysis of Customer Attrition Prediction using Machine Learning Techniques.

Akshara sri L[1], Harini Murugan[2], Nithisshkrishna KS[3],Anitha R[4]
aksharasri0473@gmail.com[1],harinimurugan@gmail.com[2],nithissh683@gmail.com[3],anitha.r@rajalakshmi.edu.in
Department of Artificial Intelligence and Machine Learning, Rajalakshmi Engineering College, Chennai, India.

Abstract — The quick development of technological infrastructure has significantly altered how organizations carry on with work. Subscription-based services are among the results of continuous digitization and customer attrition has become a major problem and a threat to all firms. Customer attrition, alternatively referred to as customer turnover, refers to the departure of customers over time, which is facing challenges in various business industries. To increase the customer retention percentage and for the overall profitability of the industry, customer churn must also be reduced. When organizations recognize client attrition, they can take proactive measures to keep customers. Customer attrition is a terminology adopted by different organizations to encapsulate the defection of clients or subscribers to any phenomenon. With the use of big data architecture, notably Spark, this study presents a web application for extracting telecom data. It uses machine learning algorithms like Logistic regression and K-means clustering, evaluates the performance of the models, and combines hard and soft data in order to predict customer churn more accurately. In addition, label selection will be carried out by assessing each feature's impurity score independently, and cluster classification will be carried out to select the best cluster based on its metrics. The study concentrates on the crucial machine learning methods for calculating client churn. This can include improving customer service, offering loyalty programs, or adjusting pricing strategies.

Keywords — Customer attrition - Apache Spark - K-Means Clustering - web application - customer retention- Logistic regression - machine learning algorithms.

I. Introduction

Customers represent the utmost value in any industry resource since they are the primary engine of profit generation. Organizations in today's world understand that they should invest in a lot of strategies that encourage customer retention and satisfaction. Churners are those who relocate to other companies for several reasons. To reduce customer turnover, the organization should possess the ability to accurately forecast the customer actions and underlying causes under their control. The binary classification task of prediction distinguishes churners from non-churners. Customer attrition prediction has become a critical area of focus which gained interest in recent times. In the telecommunications sector, competition for customers' attention is fierce. There has been a dramatic shift in the telecoms sector ever since customer attrition forecasting was introduced. The churn rate is very high in the telecom sector [9]. Businesses can potentially provide better service to customers and differentiate themselves from rivals by using input from many customer touchpoints. Many firms lose customers because they aren't able to anticipate when those customers will leave. The study's major objective is to aid telecommunications service providers in better preparing for the loss of customers. However, customer attrition is still a huge issue in the telecommunications industry. Data-driven insights for churn prediction and resource allocation have become more important for telecom businesses as they strive to retain customers in the face of fierce competition.[3]. "The rate at which a company loses customers to a competitor" is one definition of customer churn [12]. companies for several reasons. To reduce customer turnover, the organization should possess the ability to accurately forecast the customer actions and underlying causes under their control. The binary classification task of prediction distinguishes churners from non-churners. Customer attrition prediction has become a critical area of focus which gained interest in recent times. In the telecommunications sector, competition for customers' attention is fierce. There has been a dramatic shift in the telecoms sector ever since customer attrition forecasting was introduced. The churn rate is very high in the telecom sector [9]. Businesses can potentially provide better service to customers and differentiate themselves from rivals by using input from many customer touchpoints. Many firms lose customers because they aren't able to anticipate when those customers will leave. The study's major objective is to aid telecommunications service providers in better preparing for the loss of customers. However, customer attrition is still a huge issue in the telecommunications industry. Data-driven insights for churn prediction and resource allocation have become more important for telecoms businesses as they strive to retain customers in the face of fierce competition.[3]. "The rate at which a company loses customers to a competitor" is one definition of customer churn [12].

Reactive and proactive customer churn management are the two main approaches, as demonstrated by Van den Poel and Burez [1]. When a business employs a reactive strategy, it holds off on ending its service connection until the client wants it. To encourage repeat business, the firm will throw in some freebies. When a business adopts a proactive strategy, it actively looks for potential buyers. The business then offers them unique incentives, in order to keep these patrons from leaving. Using iteratively learning machine learning techniques: Popular methods like k-means clustering and logistic regression can be used to foretell customer churn.
An approach for supervised learning called logistic regression can be used to model the likelihood of a binary outcome, such

as a customer's likelihood of churning. The primary goal of this research is to help telcos better anticipate and prepare for customer defections. However, client churn remains a major problem in the telecommunications sector. In order to retain consumers in the face of intense competition, telecoms companies have come to realize the value of using data-driven insights for churn prediction and resource allocation.[3]. One definition of customer churn is "the rate at which a company loses customers to a competitor"[12].

I. METHODOLOGY

a. Random Forest Algorithm (Xiancheng Xiahou):

b. The Random Forest algorithm is used as the methodology in this study. This algorithm was used since it is an effective and reliable tool for selecting features. In many fields, including business, economics, finance, and biology, Random Forest has been adopted because of its superior classification accuracy, robustness against noise and anomalies, and capacity to generalize results. Given the dataset's considerable dimensionality of 17 variables, the challenge was to determine the number of features (M) to include in the predictive model. To address this, the Out-of-Bag (OOB) error was utilized as a standard of measurement for feature selectionTo determine the OOB error, the Random Forest was constructed with each tree using a different set of bootstrap samples from the training set, and the results changed as the number of randomly selected features grew. This suggested that the choice of the feature count (M) did not significantly impact the model's performance. Consequently, the decision was made to select four features in each iteration, resulting in a relatively low OOB error.This suggested that the choice of the feature count (M) did not significantly impact the model's performance. Consequently, the decision was made to select four features in each iteration, resulting in a relatively low OOB error. In order to foresee client attrition, it is necessary to consider four metrics: "Night Buy," "PM Buy," "Night PV," and "PM PV." These were deemed the most important determinants of client churn in the churn prediction model[15].

a. **U-Net (Karan Jakhar et al.):**

The sheer volume of information currently available makes classification into distinct types—for example, audio, video, and written content—a commonplace practice in the field of data analysis. This characterization is fundamental for viable information mining, which envelops a scope of functionalities like grouping, segregation, affiliation, and bunching. Numerous complete frameworks are intended to give a set-up of information mining functionalities inside a solitary stage (Neha and Vikram, 2015). One notable classification technique is the Support Vector Machine (SVM), which excels in handling linear permutations of subsets within a training dataset. SVM aims to find a maximum margin separating hyperplanes in a high-dimensional feature space, which is particularly useful when dealing with nonlinearly separable information highlights (Nadeem, Umar, and Shahzad, 2018). This method effectively organizes data based on the most significant characteristics, even in scenarios where the vectors are nonlinearly separable[14].In the SVM system, a few key parts assume essential parts:

M: Represents the number of samples in the training dataset.
Xi: Denotes vector support when the value of ai exceeds 0.
'X': Represents an unidentified vector sample.
δ (delta): Serves as a threshold or margin.
(ai): Parameter derived from solving a convex quadratic programming problem related to linear constraints.

The kernel functions used to transform data into higher-dimensional spaces for more precise class separation include the Polynomial kernel and the Gaussian radial basis function (RBF), to name just two examples. The threshold () is another parameter that must satisfy the Karush-Kuhn-Tucker criterion (Burges, 1998), and it is chosen by picking any 'i' where ai is bigger than 0.

In summary, SVM is a powerful classification technique that maximizes the margin between data points in a high-dimensional space. It is an important technique in data mining and classification tasks since it is especially helpful when working with complex and nonlinearly separable data.

b. **KNN (Prabadevi. B):**

Random gradient boosting is the term for this type of boosting. A sample of the training dataset is selected at random (without replacement) and used for subsequent iterations. Then, the randomly picked subsample is used to fit the base student instead of learning from the whole example. Some potential stochastic variations include: After subsampling the columns, make the individual trees. The segments should be subsample tested before each tree is made.

Rule of training.

1. The initial phase of the teaching for training.
2. Inputs
3. A tendency
4. The NN's biases and weights, as well as the learning rate, should both be set to zero for optimal performance.
5. Begin each information unit with the following Si(i=1 to n) = xi
6. After the result, get feedback from the web.
7. Using the appropriate activation function, ascertain a conclusive result based on the results of step 6.

In this research, we present stochastic gradient boosting, an improvement on the gradient boosting strategy that allows the method to be applied to categorical target variables in addition to continuous objective variables (as a classifier or regression). To reduce the model's incorrect bias, gradient boosting is employed. In the context of classification, the cost is expressed as a log loss, while in the context of regression, it is expressed as a mean squared error (MSE). Because of its adaptability,

stochastic gradient boosting can be used to enhance a broad variety of inherently unfortunate activities; furthermore, it provides a number of hyper boundary modifying options that render the capacity fit highly malleable. Data cleansing is not required [7] because the statistics and classifications can be used in their current form.

c. Churn Prediction Using Naive Bayes (Khulood Ebrah):

Naive Bayes uses the Bayes rule and a set of conditional independence assumptions to perform classification [11]. Computing P(X|Ci)P(Ci) will tell us to which category Ci X belongs.If tuple X is a member of class Ci, then the classifier will return Ci as the class label.

$$P(X|Ci)P(Ci) > P(X|Cj)P(Cj) \text{ for } 1 \le j \le m, j \ne i$$

Stated differently, the class Ci with the highest P(X|Ci)P(Ci) is the projected class label [12]. models posterior probability using the Bayesian approach. Specifically, for every k=1,⋯, K,

$$\hat{P}(Y=k|X1,\cdots, Xp) = \pi(Y=k)\pi P(Xj|Y=k) / \sum \pi(Y=k)P(Xj|Y=k)$$

where Y is an arbitrary number that represents the index of the churn class for a given observation. Predictors of an observation are the variables X1 and Xp. (Y=k) expresses the prior probability that a category index is k.The model classifies predictors based on their mean and standard deviation.

The method uses Naive Bayes classification to estimate the parameters of a probability distribution under the assumption that predictors are conditionally [4] independent given the class. First, you have to assume The posterior probability of a sample belonging to each class can be calculated in this way for data that has not yet been observed during testing. After that, the posterior probabilities of the experimental results are graded.

d. Customer Churn Analysis Using LSTM-RNN Model (Nagaraju Jajam):

In order to reliably predict the chance of customer turnover from the input dataset, the churn classification technique employs the LSTM-RNN model. To do this, we employ a deep learning architecture equipped with an attention layer to boost the accuracy of churn classification. The proposed LSTM-RNN model requires a few extra processing steps before it can be put into action. Before feeding data into an LSTM-RNN architecture, features are convolved. In this step, we focus on deciphering the meaning of the text by analyzing the sequence of its words. Additionally, the LSTM-RNN architecture effectively identifies and captures temporal correlations between features, resulting in a feature vector that contributes in the overall churn classification process [5].

Assuming the semantic meaning of input data also entails comprehending the context and underlying information contained in the data, especially as it relates to customer behavior. In order to train such a model, it is necessary to create labels that indicate whether a client has churned (assigned a value of 1) or not (given a value of 0). It is recognized, nevertheless, that these churn label assignments may be somewhat arbitrary, given that churn is frequently determined by a variety of variables and interpretations.

An LSTM-RNN is trained by first building a matrix layer from the labeled data. In specifically, a Long Short-Term Memory (LSTM) cell is a type of RNN cell that is particularly good at processing sequential input and detecting remote dependencies. There are certain drawbacks to using them, despite the fact that they perform quite well in areas like time-series analysis and natural language processing. LSTM cells are more expensive to train and run on a computational level because to their complex architecture with many parameters and actions. Despite these drawbacks, LSTM-RNNs are favored due to their higher capacity to represent sequential data.

While less processing power is required for simpler RNN cells, they aren't very good at finding patterns in data. RNNs have their own advantages inside the larger neural network framework. Thanks to their architecture based on deep neural networks, they can do sequential and concurrent data analysis, increasing their usefulness. Adding memory cells, as in long short-term memory recurrent neural networks (LSTM-RNNs), allows the network to replicate its processing abilities, especially those associated to retaining and applying information across longer sequences. For challenges like churn prediction that necessitate an appreciation for sequential data and long-term interdependence, LSTM cells may prove useful, despite their high processing requirements.

e. Automated Pneumothorax Detection and Quantification from CT Scans (Soumi De):

The Sampling-based Stack Framework (SS-IL) that has been proposed provides a new method for churn prediction. This framework makes use of ensemble learning to improve classifier performance. The outputs of several base classifiers are combined using the potent technique of ensemble learning to arrive at a final classification. Stacking is a sort of ensemble learning that employs multiple level-0 learners (also known as base learners) to learn from the same training dataset.

The SS-IL framework is unique in that it uses different training datasets for the classifiers at level 0 of the classification. By using sampling techniques, the goal is to increase the variety of attributes taken into account and make it easier for the ensemble to gather important information. The goal of this training data diversification is to raise the framework's overall predictive power.

In addition, the predictions made by the level-0 learners are used to train a meta-learner, another part of the SS-IL system. Effective instance classification is made possible by this meta-learner's acquisition of the combination weights for each of the decision probabilities supplied by the base-level classifiers. A stacked ensemble such as SS-IL relies on the

level-0 base learners to facilitate the information gained from the features used in training the meta-learner. This framework is essentially based on the idea that multiple classifiers' combined knowledge and the variety of training data improve predictive robustness and accuracy.

Although the SS-IL architecture is addressed in this article in the context of churn prediction, there are hints that it may also have medical uses, particularly in the fields of pneumothorax monitoring and diagnosis. The framework highlights its significance beyond predictive analytics by demonstrating its versatility and utility across different domains, potentially improving patient care and saving time.

III. Conclusion

Business research in the telecommunications industry is conducted with the intention of boosting the industry's bottom line. Predicting customer attrition is the telecom industry's principal source of income. In this paper, we look at how to use big data to create an app that can accurately forecast client churn. Predicting customer churn with big data and machine learning is a huge boon to the telecom industry. We were aware of the significance of data cleansing and preparation for analysis. To determine what factors most significantly affect customer churn, we set out to examine pertinent features and patronage patterns. Using methods like logistic regression and K-means clustering, crucial characteristics may be extracted from mountains of telecom data and fed into machine learning algorithms that anticipate and prevent customer turnover. The reviewed methods show potential for creating a comprehensive machine-learning model to benefit the industry in retaining customers. These tactics have the potential to significantly improve the industry's bottom line and the delight of its customers. Telecom companies can improve customer retention, customer satisfaction, and revenue loss due to churn by precisely identifying consumers at risk of leaving and tailoring retention activities to meet the unique needs of each customer. Our study's overarching objective is to help the big data analytics and telecommunications communities better understand customer attrition prediction and to inspire new lines of inquiry and fresh ideas.

IV. REFERENCES

1. Burez J., & Van den Poel, D “Crm at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services”, Expert Systems with Applications 32, 277–288.
2. Ledro, C., Nosella, A., & Vinelli, A. (2022). Artificial intelligence in customer relationship management: literature review and future research directions. Journal of Business & Industrial Marketing, 37(13), 48-63.
3. Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn prediction in telecommunication using logistic regression and logit boost. Procedia Computer Science, 167, 101-112.
4. Khulood Ebrah, Selma Elnasir “Churn Prediction Using Machine Learning and Recommendations Plans for Telecoms”.Journal of Computer and Communications > Vol.7 No.11, November 2019.
5. “Arithmetic Optimization With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction” in IEEE vol 11.
6. Soumi De, Prabu.P” A Sampling-Based Stack Framework for Imbalanced Learning in Churn Prediction in IEEE vol 10.
7. Prabadevi.B, Shalini.R, Kavitha.B.R (2023). Customer Churning analysis using machine learning algorithms. In International Journal of Intelligent Networks.
8. M. Alizadeh, D. S. Zadeh, B. Moshiri and A. Montazeri, "Development of a Customer Churn Model for Banking Industry Based on Hard and Soft Data Fusion," in IEEE Access, vol. 11, pp. 29759-29768, 2023, doi: 10.1109/ACCESS.2023.3257352
9. Anand, M., Shaukat, I., Kaler, H., Narula, J., & Rana, P. S. Hybrid Model for the Customer Churn Prediction
10. Zadoo, A., Jagtap, T., Khule, N., Kedari, A., & Khedkar, S. (2022, May). A review on churn prediction and customer segmentation using machine learning. In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) (Vol. 1, pp. 174-178). IEEE.
11. Mitchell, T.M. (2015) Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression.
12. Han, J., Pei, J. and Kamber, M. (2011) Data Mining: Concepts and Techniques. Elsevier, Amsterdam.
13. PM, U., & Balaji, N. V. (2019). Analyzing Employee attrition using machine learning. Karpagam Journal of Computer Science, 13, 277-282
14. Abdulsalam Sulaiman Olaniyi , Arowolo Micheal Olaolu , Bilkisu Jimada- Ojuolape , Saheed Yakub Kayode,,” Customer Churn Prediction in Banking Industry Using K-Means and Support Vector Machine Algorithm. In International Journal of Multidisciplinary Sciences and Advanced Technology Vol 1 No 1 (2020) 48–54.
15. Xiancheng Xiahou and Yoshio Harada , "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM
16. Babu, R., Jayashree Kannappan, Brahmadesam Viswanathan Krishna, and K. Vijay. "An efficient spam detector model for accurate categorization of spam tweets using quantum chaotic optimization-based stacked recurrent network." Nonlinear Dynamics (2023): 1-18.
17. Kandasamy, Vijay, Revathy Padmanabhan, Priya Vallinayagam, and Sowmia Kanakam Rajendran. "Survey on chaos RNN–A root cause analysis and anomaly detection." In AIP Conference Proceedings, vol. 2790, no. 1. AIP Publishing, 2023.