e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

CUSTOMER CHURN PREDICTION

Kanagaraj S[*1], Swathy K[*2], Shalini S[*3], Selvakumar[*4]

[*1,2,3,4]Department Of Information Technology, Kumaraguru College Of Technology,
Coimbatore, Tamil Nadu, India.

DOI : https://www.doi.org/10.56726/IRJMETS32736

ABSTRACT

The concept of Churn has been used for years to increase profitability and stabilize customers-company relationships. So this is becomes the major issue in business industry in multiple domains such as Retail, Telecom industry, Finance, Online Music Streaming, Online Game, Internet Funds, Insurance etc. The motive is to predict whether the users will churn or not. There are some hypothesis to find whether the activities of the customer. In this paper we are using Sparkify dataset to predict the churn activities. Sparkify is a Online music Streaming application which is similar to Spotify. In this paper we have compared the important features using Particle Swarm Optimisation(PSO). Here we have compared many algorithms like, Deep Neural Network from Deep Learning, Light Gradient Boosting, Bagging from Machine Learning Algorithms along with some of the metrics like Senstivity, Specificity, Accuracy, Precision etc.

Keywords: Churn, Online Music Streaming, Particle Swarm Optimization(PSO), Machine Learning, Deep Learning.

I. INTRODUCTION

Churn prediction is a process of predicting which customers will be at high risk of canceling a subscription or leaving the network, based on the customer's behavior based on the product. It is a major critical issue for many business because gaining new clients every time, it costs more than maintaining the existing ones. Customer's maintainenace is one of the major objectives of Customer's Relation Management, customer churn occurs when the user terminates or leaves from a service or company. The major issue of churn is due to the discontent of the service given by the provider. There are many different factors which cause churn and it also differ for each customer. Many business employee many techniques to predict churn and grow many ideas to retain customers to longer tenure. Customer's prediction has enhance the number of one business goal. In general, deep learning techniques and machine learning techniques are employed to perform customer churn prediction with better accuracy.

TYPES:

Churners are mainly classified into two types based on categories as Voluntaries and Involuntaries. Voluntary churn occurs when the customer begins to interrupt service, and is further divided into intentional and accidental churners. Involuntary confusion occurs when the company starts removing customers from the subscriber's list. Secondary churners occur due to the incident because of few changes in the location or change in the financial position and deliberate churners occur, due to some customers need to change the technology or price rate.

II. RELATED WORKS

Gaddam, Lahari, Kadali, Sree Lakshmi Hiranmayee et al (2022) [1] they research about customer churn prediction done in online music streaming. It mainly focuses on two main objectives, the first one includes the literature review of the latest research work predicted in churn for music streaming services. Secondly, it compares the performance and accuracy of four supervised machine learning techniques, to find which algorithm performs and best suited for churn prediction with 89% accuracy.

Martins, Helder et al (2017) [2] the providers of online streaming services have witnessed a fast growth of the customer's base in last few years. This circumstance has attracted the rapid number of competitors on obtaining their share of market. On other hand, recurrent neural networks (rnn), and the long short-term memory (LSTM) variant, have shown good results for the other domains like Insurance, Telecom, Game, Music, Fund etc, where data is treated as sequence instead. They shown the best acuuracy of 88.8%.

Wael Fujo, Samah; Subramanian, Suresh; and Ahmad Khder, Moaiad et al (2022) [3] To solve the unbalanced issue, Random Oversampling technique used to evaluate both datasets. The results show the model implemented and performs well with the lasso regression for the feature selection, early stopping of the technique to pick epochs, and the large number of neurons to the input and the hidden layers, and activity regularization that minimize overfitting for the both datasets. In predicting the customers churn, our outperform Machine Learning techniques: XG_Boost, Naïve_Bayes, Logistic_Regression and KNN. Moreover, the Deep-BP-ANN model's accuracy that outperforms the existing deep learning techniques that will use holdout or the 10- fold CV for same datasets.

Seyed Mohammad Sina Mirabdolbaghi and Babak Amiri et al (2022) [4] They discussed various methods for reducing the features using Autoencoders, PCA, LDA, Xgboost and T-SNE. The first phase is to preprocess the given dataset so that missing values and corrupt values are handled and the data will be scaled. Second phase is to implement a complete feature reduction based on well known algorithms that reduces the features and selects the suitable one. In third phase, light GBM's hyperparameter will be tuned using Bayesian's hyperparameter optimization and the genetic optimization algorithms. It outperforms ensemble and the ML algorithms like Boosting-AdaBoost, decision tree and SVM on over the seven evaluation metrics: accuracy, Kappa ,area under the curve (AUC), Mathews correlation coefficient (MCC), F1 score, Brier score and EMPC.

Kim S, Choi D, Lee E, Rhee W (2017) et al [5] Internet-connected devices, especially mobile devices such as smartphones, have become widely accessible over the past decade. Interaction with such devices has evolved into frequent and short-term use, and this phenomenon has resulted in the widespread popularity of casual games in the sports industry. On the other hand, the development of casual games has become easier than ever as a result of the advancement of development tools. As a result of fierce competition, both acquisition and retention of users are now prime concerns in the industry. Overall results indicate that while the analysis results provide useful insights, a small number of well-chosen features used as performance metrics are sufficient to make important operational decisions, and OP and CP should be properly selected depending on the analysis goal.

Luo Bin, Shao Peiji, Liu Juan (2007) et al [6] In today's fast-changing and highly competitive telecommunications market, forecasting and management are critical for more and more companies. To improve customer retention, telcos can predict at-risk customers who are likely to switch service providers. The empirical evaluation results indicate that the customer churn models constructed by sampling optical selection have good performance, and show that the proposed methods and techniques are effective and feasible when customer information is sparse and the class distribution is skewed. This study benefits not only predictive research and practice, but also other data mining applications with similar characteristics.

Lalwani, P., Mishra, M.K., Chadha, J.S. et al(2022) [7] One of the most challenging problems in the telecom industry is customer confusion (CCP). With advancements in the field of machine learning and artificial intelligence, the possibilities of predicting customer confusion have increased significantly. Our proposed method consists of six phases. In the first two phases, data preprocessing and feature analysis are performed. In the third step, feature selection is considered using a gravity search algorithm. Next, the data is divided into two parts and the test is set at the ratio of 80% and 20% respectively. Finally, the results obtained on the test set have been evaluated using the confusion matrix and the AUC curve. Adaboost and XGboost classifier gives maximum accuracy of 81.71% and 80.8% respectively. The highest AUC score of 84% is achieved by both AdaBoost and XGBoost classifiers, which outperform the others.

Qi, J., Zhang, L., Liu, Y. et al. (2009) [8] In this paper, we propose a model that combines the advantages of ADTreesLogit, ADTrees model and logistic regression model to improve the prediction accuracy and interpretability of existing churn prediction models. We show that the overall predictive accuracy of the ADTreesLogit model compares favorably with TreeNet, which won the gold prize in the 2003 Mobile Customer Churn Predictive Modeling Tournament (The Duke/NCR Teradata Churn Modeling Tournament. In fact, ADTreesLogit has better predictive accuracy than the TreeNet on two important observation points.

T. Verbraken, W. Verbeke and B. Baesens et al (2013) [9] . Interest in data mining techniques has grown tremendously in the past decades, and many classification techniques have been used in a wide range of business applications. Therefore, the need for adequate performance measures is more important than ever. In this paper, a cost-benefit analysis framework is formalized to define performance measures aligned with end-users' key objectives, i.e., profit maximization. The advantage of this approach is that it helps companies in choosing the classifier that maximizes profits. Also, it helps operationalization by providing guidance on the segment of the customer base that should be included in the retention campaign.

K. Dahiya and S. Bhatia et al(2015) [10] With the rapid growth of the telecom industry, service providers are leaning more towards expanding their subscriber base. Retaining the existing customers has become the biggest challenge to meet the demand to survive in the competitive environment. A study conducted in the telecom industry says that the cost of acquiring a new customer is more than retaining an existing customer. Our paper proposes a new structure predictive model and implements it using WEKA data mining software. The effectiveness and efficiency of decision tree and logistic regression techniques were compared.

III. PROPOSED SYSTEM

The contribution of the proposed research is to design a machine-learning- and deep learning based intelligent decision support system for the churn prediction. Here we use Light Gradient Boosting it is a kind of Boosting Technique and bagging Techniques from machine learning algorithms and deep neural network from deep learning for Online Music Streaming. Still there are only existing solutions for other domains not for Music streaming. In this project we use PSO for feature extraction and k fold cross validation method is used for splitting of training and testing data

And we add new feature called “CHURN” by analyzing all other features the various machine learning and deep learning techniques are used and finally metrics are evaluated in various kinds using those metrics we will compared and analysed the output among various techniques that which will give the best accuracy. Finally we use Grafana tool for visualizing the dataset.

IV. METHODS

DATA COLLECTION:

It may contains a complete value or some missing values in some fields.we collected our mini spark dataset from browser and it contains 17000 rows and 18 columns in it.

DATA PREPROCESSING:

Data pre-processing involves several procedures. A meaningful and valuable information will be extracted and completed from the original dataset. Selection Features are essential in determining relevance properties from a dataset using domain knowledge. The dataset used in this research consists of 16000 records with 18 characteristics. There are many missing items in the Churn dataset, invalid values like "null" and unbalanced features. As a preprocessing step, we removed many outliers and missings in a filter that reduces values and attributes counts only to useful ones. In this study, for appropriate feature selection, we used a particle swarm Optimization Algorithm. The extracted feature is used input to classification algorithms.

FEATURE SELECTION (PARTICLE SWARM OPTIMISATION):

PSO is an evolutionary optimization approach based on swarm dynamics and intelligence. We used PSO to select relevant features that have a significant impact on the performance of evolution. It calculates the minimum value of the function. Each particle tries to change its position x by using the formula:

$$x\,(t+1) = x\,(t) + v\,(t+1)$$

$$v\,(t+1) = Wv\,(t) + c1 \times rand\,(\,) \times (\,Xpbest - X\,(t)) + c2 \times rand\,(\,) \times (\,Xgbest - X\,(t))$$

v (t) is the particle velocity at time t, and v (t) is the particle position at time t. inertia weight W, c1, c2 is a learning factor, rand is a random integer between zero and one, the ideal particle state is Xpbest, while the ideal global state is Xgbest. Each individual in PSO is referred to as a particle. It is characterized as a viable solution to the optimization problem in D-dimensional search space, with the ability to recall swarm and best locations and speed.

ALGORITHMS USED:

BAGGING:

A Bagging classifier is a ensemble meta-estimator that fits the base classifiers each on the random subsets of the given original dataset and then it aggregate their individual prediction (either by voting or by averaging) to form the final prediction. Such an meta-estimator can be typically used as a way to the reduce the variance of the black-box estimator (e.g., a decision tree), by introducing the randomization into its construction of the procedure and then making an ensemble model out of it. Each base classifier is trained in parallel with a training set, which is generated by randomly drawing N examples (or data) from the original training dataset - where N is the size of the original training set.The training set for each base classifier is independent of each other. Many of the original data may be repeated in the resulting training set, while others may be omitted. Packing reduces overfitting (variance) by averaging or voting, however, this leads to an increase in bias, although this is offset by a reduction in variance. Because packing modifies the original training dataset, some events (or data) may exist multiple times, while others may be missing.

LIGHT GRADIENT BOOSTING:

LightGBM is a gradient boosting framework based on decision trees to increase model performance and reduce memory usage. It uses two innovative techniques: a gradient-based one-sided model and an exclusive feature set (EFB) that overcomes the limitations of the histogram-based algorithm used primarily in all GBDT (Gradient Boosting Decision Tree) frameworks. Both techniques of GOSS and EFB, described below, develop properties of the LightGBM algorithm. Together they make the model work efficiently and give it a sophisticated edge over other GBDT frameworks.

Gradient-based One Side Sampling Technique for LightGBM:

Different data instances have different roles in computing the information gain. Events with larger gradients (ie, events without training) will contribute more to information acquisition. GOSS keeps those events with large slopes (eg, larger than a predefined threshold, or in the upper percentiles), and randomly discards only those events with small slopes to maintain the accuracy of the information gain estimate. This treatment can lead to more accurate gain estimation than uniform random sampling with the same sampling rate, especially when the value of the information gain has a large range.

DEEP NEURAL NETWORKS:

A deep neural network is an artificial neural network with the multiple hidden layers between input and output layers. Like shallow ANNs, DNNs can model complex nonlinear relationships. The main purpose of a neural network is to take a set of inputs, perform progressively complex computations and give an output to solve real-world problems such as classification. We restrict ourselves to the feed forward neural networks(FNN) . Neural networks are widely used in supervised learning (SL) and reinforcement learning (RL) problems. These networks are predicated on a set of interconnected layers. In deep learning, the number of hidden layers, often nonlinear, can be large; Say around 1000 layers. DL models produce better results than ordinary ML networks. We often use the gradient descent method to optimize the network and minimize the loss function.

MEASURE INDEX:

To obtain the most reliable results from a scientific investigation, it is crucial to avoid bias and error while being accurate and precise in data collection. Accuracy and precision are both concerned with how close a measurement is to its true or true value.

$$\mathbf{ACCURACY} = \frac{\mathbf{True\ Positive\ (TP) + True\ negative (TN)}}{\mathbf{ALL}}$$

Precision (also known as positive predictive value) is the fraction of relevant events among retrieved events, while recall (also known as sensitivity) is the fraction of relevant events retrieved. So both precision and recall are based on relevance.

$$\mathbf{PRECISION} = \frac{\mathbf{True\ Positive\ (TP)}}{\mathbf{True\ Positive\ (TP) + False\ Positive (FP)}}$$

Recall is a metric that counts the number of correct positive predictions made out of all positive predictions that could have been made.

$$\mathbf{RECALL} = \frac{\mathbf{True\ Positive\ \ (TP)}}{\mathbf{True\ Positive\ (TP) + False\ Negative\ (FN)}}$$

Sensitivity (true positive rate) refers to the probability of a positive test being a true positive.

IRJMETS

e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

$$\mathbf{SENSITIVITY} = \left(\frac{\mathbf{True\ Positive\ \ (TP)}}{\mathbf{True\ Positive\ (TP) + False\ Negative\ (FN)}}\right) \times 100$$

Specificity (true negative rate) refers to the probability of a negative test being truly negative.

$$\mathbf{SPECIFICITY} = \left(\frac{\mathbf{True\ Negative\ (TN)}}{\mathbf{True\ Negative\ (TN) + False\ Positive\ (FP)}}\right) \times 10\mathbf{0}$$

The F1-score metric that uses the combination of both precision and recall. The F1 score is an harmonic mean of two.

$$\mathbf{F1\ Score} = \frac{\mathbf{2}}{\left(\frac{\mathbf{1}}{\mathbf{PRECISION}}\right) + \left(\frac{\mathbf{1}}{\mathbf{RECALL}}\right)}$$

V. SYSTEM REQUIREMENTS

Language: Python (Pyspark)

IDE: Anaconda, Notebook

Framework: Apache Spark

Visualization Tool: Grafana

VI. CONCLUSION

Results are achieved by machine learning and deep learning algorithm and feature selection techniques are presented in this section. For selecting features, we use PSO(Particle Swarm Optimisation) and then Machine learning algorithms and Deep learning techniques with cross-validation technique. For the performance measures, we used precision, accuracy, recall, Area under curve,F1 score, Confusion matrix,

Sensitivity, Specificity using these metrics we find which algorithms produces high accuracy for this prediction according to these metrics.

VII. REFERENCES

[1] Gaddam, L., & Kadali, S. L. H. (2022). Comparison of Machine learning algorithms on Predicting Churn within Music streaming service (Dissertation).

[2] Martins, H. (2017). Predicting user churn on streaming services using recurrent neural networks (Dissertation).

[3] Wael Fujo, Samah; Subramanian, Suresh; and Ahmad Khder, Moaiad (2022) "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," Information Sciences Letters: Vol. 11 : Iss. 1 , PP -.

[4] Seyed Mohammad Sina Mirabdolbaghi, Babak Amiri, "Model Optimization Analysis of Customer Churn Prediction Using Machine Learning Algorithms with Focus on Feature Reductions", Discrete Dynamics in Nature and Society, vol. 2022, Article ID 5134356, 20 pages, 2022.

[5] Kim S, Choi D, Lee E, Rhee W (2017) Churn prediction of mobile and online casual games using play log data. PLoS ONE 12(7): e0180735

[6] L. Bin, L. Juan, " Churn prediction based on decision tree in personal mobile telephone system service," 2007 in International Conference on Service Systems and Service Management(ICSS), 2007, pp. 1-5,

[7] Lalwani, P., Mishra, M.K., Chadha, J.S. et al. Customer churn prediction system: based on a machine learning approach. Computing 104, 271–294 (2022).

[8] Qi, J., Zhang, L., Liu, Y. et al. ADTreesLogit model for customer churn prediction. Ann Oper Res 168, 247–265 (2009).

[9] T. Verbraken, B. Baesens and W. Verbeke, "A Novel Profit for Maximizing the metrics for Measuring Classification Performance of the Customer Churn Prediction Models," in IEEE Transactions on the Knowledge and Data Engineering, vol. 25, no. 5, pp. 961-973, May 2013,

[10] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," 2015 4th International Conference on Reliability(ICR), Infocom Technologies and Optimization (ITO) (Trends and Future Directions), 2015, pp. 1-6