International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-13 Issue-1, May 2024

A Review on Classification Algorithm for Customer Churn Classification

Nurul Nadzirah Adnan, Mohd Khalid Awang

Abstract*: Any sector faces a huge obstacle when it comes to retaining existing customers. The percentage of consumers who have quit using a product or service is referred to as customer churn, and it is a vital indication that offers reliable information about this percentage. When it comes to achieving long-term success in a market or industry, one of the most significant challenges that any company must face is the ability to keep their precious clients and to fulfill their needs. A review of the most significant studies on Customer Churn Prediction is presented in this paper so as to furnish the reader with an overview of frequently employed data mining methodologies and their respective performances. We provide the available statistics in addition to customer information in order to approximate customer attrition. The time period encompassing the survey extends from 2003 to 2023. During the process of Customer Churn Prediction, we identified the issues and difficulties that were linked with it and offered guidance and potential remedies.

***Keywords*: Artificial Neural Network, Churn Prediction, Survey, Customer**

I. INTRODUCTION

This paper offers a comprehensive overview of churn and churn prediction, including its impact on sectors and the underlying causes of churn. The study examined various techniques employed for churn prediction in the existing body of research. Furthermore, the current world is witnessing rapid expansion in a variety of fields, particularly technology. The increasing demand is accompanied by a rising sense of competitiveness among different brands and rivals. Customers are a valuable resource for every business. Acquiring a thorough comprehension of the consumer demographic and their behavioral patterns is quite significant. In order to address this problem, it is crucial that we focus our attention on client segmentation. Customer segmentation enables the focused identification and targeting of customers to improve personalized user experiences, ultimately leading to increased customer loyalty. Churn refers to the action taken by a client to change service providers, typically due to dissatisfaction with the services provided [1]. To mitigate churn, service providers must anticipate client attrition by analyzing their behavioural patterns. The continuous process of churning has a direct impact on the overall business profitability and reputation. Therefore, it is always preferable to predict and preempt customer churn. The examination of client record information is currently experiencing a surge in popularity, especially within the telecommunications, retail, and banking sectors, owing to its considerable significance. Due to the fact that the cost of acquiring new consumers is greater than the cost of retaining existing ones, prediction is vital. Hence, even marginal enhancements and progress in the attrition prediction model have the potential to generate substantial financial expansion for businesses. An exhaustive analysis of customer attrition prediction studies conducted from 2003 to 2023 is presented in this article. interest in developing a loss prediction model for various industries has remained constant. You must employ data analysis techniques such as clustering, pattern recognition, extraction, pre-processing, and classification in order to predict when a customer will depart. To implement these techniques, conventional classifiers, ensemble classifiers, and other hybrid methods are utilized. This essay provides a comprehensive analysis of several machine learning algorithm models utilized in the business world for loss prediction. The articles are systematically analyzed and categorized by taking into account the characteristics, methodologies, and machine learning techniques employed. The introduction of ensemble and hybrid techniques has resulted in an increase in the accuracy of predictive models. The organization of this paper is as follows: In section 2, we examined the process of article selection using the Systemic Analysis Procedure for Electing Articles. Section 3 provides a comprehensive classification of articles. In Section 4, a variety of data sources used for Customer Churn Prediction are presented. Section 5 discusses the limitations, challenges, future research and methods used for feature selection in Churn Prediction. Finally, Section 5 serves as the concluding part of this article.

II. SYSTEMATIC ANALYSIS PROCEDURE FOR ELECTRIC ARTICLES

The research articles in this paper are gathered and chosen based on the Systemic Analysis Procedure (SAP). This strategy assists in selecting the optimal articles to address research inquiries in a proficient and suitable manner. At first, we collected 951 articles pertaining to research inquiries. Subsequently, we eliminated 476 papers on the grounds of having irrelevant abstracts and content that fell outside the intended scope. The duplication phase in 205 eliminated papers. In addition, 217 papers have been excluded during the reviewer phase due to their substandard quality.

Manuscript received on 11 March 2024 | Revised Manuscript received on 27 March 2024 | Manuscript Accepted on 15 May 2024 | Manuscript published on 30 May 2024.

*Correspondence Author(s)
Nurul Nadzirah Bt Adnan*, Department of Informatic & Computing, University Sultan Zainal Abidin, Terengganu, Malaysia. Email: nurulnadzirahadnan@gmail.com, ORCID ID: 0009-0003-6225-708X
Mohd Khalid Bin Awang, Department of Informatic & Computing University Sultan Zainal Abidin, Terengganu, Malaysia. Email: khalid@unisza.edu.my

A. Research Questions:

Research Queries consists of three categories of questions: (a) Inquiries pertaining to the application of Machine Learning techniques in Customer churn Prediction, (b) Questions regarding customer churn datasets, and (c) Inquiries concerning future trends and opportunities. Table 1 depicts the research questions for Customer churn prediction.

Table 1. Research Question for Customer Churn

S. No	Questions
RQ1	What is customer churn?
RQ2	Which machine learning algorithm is used for classification, clustering, and optimizing customer churn classification?
RQ3	What are the primary machine learning approaches utilized in predicting customer churn?
RQ4	What are the classifications of public and private datasets utilized in churn prediction? What is the regularity of usage for these datasets?
RQ5	What is the process for combining a single classifier into the creation of multiple classifiers?
RQ6	What is the definition of multi-level stacking ensembles? What factors contribute to its current widespread popularity?
RQ7	What is the significance of churn prediction for providers? The impact of client attrition on service providers

B. Articles Source:

The papers are gathered from reputable sources specified below, covering the time period from 2013 to 2023.

- IEEE Explorer
- Elsevier
- Springer
- Google Scholar
- ACM Digital Library

C. Search Phrase:

- Customer Churn Prediction
- Churn Classification
- Customer retention
- Churn prediction

D. Inclusion and Exclusion Aspects:

- Articles must be sourced from reputable publishers and be available for download.
- Articles specifically focused on the application of Telecom industry.
- Articles must contain high-quality content that is directly related to the topics of binary classification, clustering, prediction, and identification of churners.
- It is required to present ideas or solutions to address the problems and issues related to customer churn in the telecommunications industry.
- Articles must be pertinent to the field of machine learning and its optimization algorithms.
- Papers must not be a literature review or an examination paper.
- The articles are in languages other than English.
- Papers that contain duplicated content, lack effectiveness, and have not undergone peer review.

III. PROBLEM DEFINITION

One of the first problems with current study is that most prediction models only use a few classifiers, which means they aren't very accurate [2][39]. Still, the ensemble of classifiers technique has been used in many recent studies. This method uses all the basic classifiers to get the final result. This limitation shows up when an ensemble mix has classifiers that are similar, which makes classification less accurate.

The fact that most of the modeling methods only used one classifier is another problem. Several methods used ensemble stacking, and most of them used all of the basic classifiers to make their final decisions. According to the research, ensemble stacking methods have been shown to be a good way to make the classification model work better [3]. It does, however, make a lot of big base classifiers, some of which are useless and similar, which makes the classification performance less than ideal. So, this study needs to include a way to find the group of ensemble stacking classifiers that works best. The present study also has a limitation in that it usually includes all the known data about customer churn in the dataset. It has been agreed that this limitation exists because not all features would have a big effect on predicting customer churn [3]. In fact, some traits may not be useful and, as a result, get in the way of learning.

IV. TAXONOMY OF ELECTED ARTICLES

The articles are systematically examined and categorized according to their characteristics, methodologies, and employed machine learning techniques. The articles are put into four main groups based on these criteria which are traditional single methods, hybrid classifier methods, ensemble classifier methods, and hybrid ensemble classifiers. The classification of different types of Churn Prediction Techniques from 2003 to 2023 is shown in Figure 1.

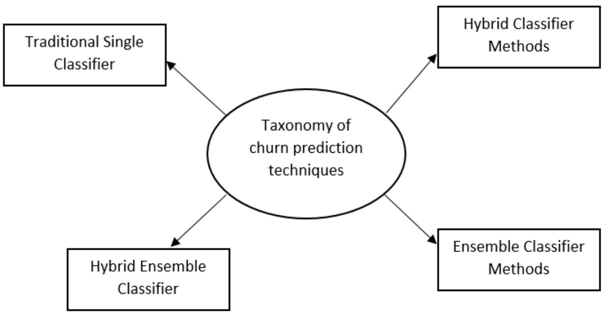

Figure 1. Classification of different Types of Churn Prediction Techniques

There are a lot of common techniques for traditional single classifier methods, like logistic regression, Support Vector Machines (SVM), and Decision Trees. On the other hand, mixed classifiers are made by putting together two or more separate classifiers. To improve precision, ensemble classifiers like boosting, stacking, and bagging are used. After using ensemble and mixed methods, the efficiency of the models has been seen to go up. Hybrid ensembles use ensemble methods to combine more than one predictor. These days, hybrid ensembles are very good at predictive data analytics and are becoming more famous because they can make better predictions.

Retrieval Number: 100.1/ijrte.A803013010524
DOI: 10.35940/ijrte.A8030.13010524
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

A. Customer Churn Classification

The word "churn" comes from the words "turnover" or "attrition." The act of people ending their relationship with a certain choice is called "churn." When a customer stops using a service, this is called "churn." Customer loss is when a customer changes their mind about what they want to buy. The customer can stop the services on their own, or the service provider can close their account. In many business areas, there are two main types of churners [5]: those who do it on purpose, also called "voluntary churners," and those who do it by accident, also called "involuntary churners." As the name suggests, intentional churners are people who choose to stop using a service on their own. Customers leaving is mostly because they are unhappy with the current plans and services. After that, these customers switch to a different service provider whose rates and services are better. People who are removed by service companies without their permission are called "unintentional churners." When this happens, the company stops providing customer service. This group of clients is the second type, which is called "unintentional" or "involuntary." Figure 2 shows how churners are put into groups.

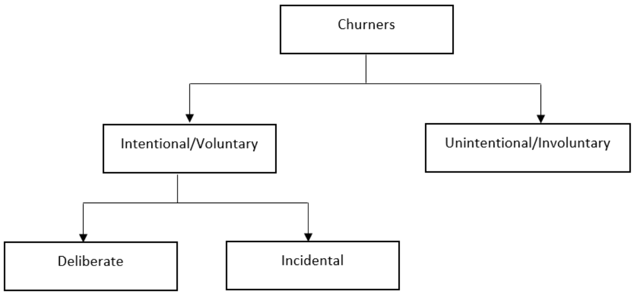

Figure 2. Categorization of Churners

Figure 2 shows how conscious churners can be broken down into two groups: deliberate churners and incidental churners. Deliberate churners are customers who are unhappy with a company's services and recharge/monthly plans and choose to switch to a different one in order to get better plans. People who cancel their membership for reasons they can't control, like moving to a place where the service provider isn't available, leaving the country, or having to use a certain service provider for work are examples of incidental churners. The churn prediction model is meant to find people who are planning to leave. By looking at how deliberate churners have behaved with the service provider in the past, different models can be used to make more accurate predictions about them. A churn prediction model is made to find people who are planning to leave, not people who are accidentally leaving. But models have a hard time correctly predicting incidental churners because customers' job and location changes are hard to predict. Some buyers might not even be aware of these changes until a certain point in time. Also, service companies don't have many ways to keep incidental churners even if they know about them ahead of time.

B. Traditional Single Classifier Methods

Traditional churn classification methods often employ distinct machine learning algorithms to forecast the likelihood of client turnover by Smith at 2000 [4]. These strategies concentrate on utilizing a solitary model to examine pertinent characteristics and generate binary forecasts concerning customer retention. Decision Trees, Support Vector Machines, Bayesian Networks, Regression, and Neural Networks are the classifiers that are commonly utilized in traditional single classifier approaches. In order to foretell which users will abandon a dataset, churn prediction models have been developed using individual classification techniques. Table 1 provides a quick overview of popular machine learning algorithms for predicting churn in traditional single classifier, highlighting their reliability and efficiency between 2003 and 2023.

Table 2. Traditional Single Classifier

Traditional Single Classifier	Definition
Logistic Regression (LR)	People who need to make two decisions, like those who need to guess how many people will leave a company, often use a linear model called logistic regression. It looks at different parts of the data to figure out how likely it is that a customer will stop using the service.
Decision Trees (DT)	Can forecast churn. The data is recursively separated by the most important attributes to create a classification tree.
Random Forest (RF)	Makes a lot of decision trees and then puts all of their predictions together. Another choice tree might not work as well or as reliably as this one.
Support Vector Machines (SVM)	Good way to divide things into two groups. The hyperplane that best splits the data into the different groups is found.
Naïve Bayes (NB)	Based on Bayes' theory, Naive Bayes is a likely classifier. It thinks that the traits are conditionally independent, which isn't always the case, but it might still work well in real life.
K-Nearest Neighbors (KNN)	KNN is a simple algorithm that sorts data points into groups based on which group their k nearest friends are in. Some situations where local trends are important can make it work well for predicting churn.
Neural Networks (Single-Layer Perceptron or Multi-Layer Perceptron)	Single-layer or multi-layer perceptrons are thought to be more modern, but they can still be seen as standard classifiers. Neural networks can figure out how features and goal variables are related in complicated ways.
Gradient Boosting Machines (e.g., XGBoost, LightGBM)	Gradient boosting algorithms, such as XGBoost and LightGBM, create a group of weak learners, which are usually decision trees, one after the other. Each tree fixes the mistakes made by the trees that came before it.
Linear Discriminant Analysis (LDA)	LDA is a way to reduce the number of dimensions in a data set that can be used for classification. It finds the linear combinations of traits that make the classes most different from each other.

It was shown by Michael C. Mozer and his colleagues in 2000 that Neural Networks and Linear Regression could be used to predict revenue loss [7]. A secret dataset from a wireless telecom company was used to test the model. In 2005, Yu Zhao, Bing Li, Xiu Li, Wenhuang Liu, and Shouju Ren [8] used a Bayesian classifier on Teradata from Duke University and got a 68% success rate. In 2005, Yu Zhao et al. published a study that used a one-class Support Vector Model to find problems in Teradata from Duke University. This model was 87.1% accurate in finding problems.

Jones et al. (2006) used K-means, Deep Neural Networks (Back Propagation), and Decision Tree (C5.0) in their study [9]. Using data from a Taiwanese telecom company over a year, their study tried to make prediction models and divide consumers into groups. Hit ratio and Lift measures were used to judge how well their models worked. In 2008, Brown et al. utilised the Support Vector Machine algorithm on the UCI churn Dataset [2]. They found that the Radial Basis Function achieved better results, with an accuracy rate of 90.9%, compared to the SVM with Radial Basis Function, which only achieved an accuracy rate of 59%. Clark et al. (2008) [5] employed a Bayesian Belief Network to discern efficient churn control tactics by analysing customer behaviours. They applied their model to a dataset from a Turkish telecommunications company. The CHAID approach was utilised to discretize continuous variables. Logistic Regression was used on a private dataset by Marcin Owczarczuk et al. to look at loss models for both prepaid and post-paid customers. Using the lift curve, they judged the models and suggested areas for future study. In their study, Thompson et al [6]. created a profit metric and performed trials using multiple classification algorithms like Logistic Regression, Decision Tree, Naïve Bayes and others on eleven datasets. The results showed that Decision Trees outperformed the other algorithms and achieved the highest performance. Gaurav Thakre in 2024 [7] shared that Miller et al. (2015) utilised six algorithms, namely ANN, LR, DT, NB, and SVM to analyse a real-life dataset from the telecommunications industry in Ireland. They incorporated a unique feature selection method and evaluated the algorithms' effectiveness by measuring the rates of true and false churn. In 2019, Irfan Ullah, Basit Raza, Ahmad Kamran Malik, et al [8]., used machine learning techniques on two datasets to test the suggested model many times. The results that were obtained using different machine learning techniques. Because it handled our data well and worked better than other methods, the random forest algorithm gave us the best results. Random Forest makes a guess by putting together a group of decision trees. Random Forest and J48 had a better F-measure score overall which is 88%. Yuyun Yulianti [9] and her research team proposes the optimal model for achieving the highest performance value is the one that incorporates the Sequential Backward Selection (SBS) and Sequential Backward Floating Selection (SBFS) feature selection techniques, specifically with 19 features. The Mutilayer Perceptrons (MLP) ensemble, trained using Negative Correlation Learning (NCL), is employed to forecast customer turnover in telecommunications firms. The study's findings validate that an ensemble of MLP models based on NCL can outperform an ensemble of MLP models without NCL, as well as standard data mining techniques commonly employed for customer churn analysis. The dataset utilized for customer churn predictions contains numerous variables, some of which may exhibit repetition or irrelevance, potentially leading to a decline in classifier performance. The goal of this work is to use feature selection methods to find and pick the most important features, which will make customer churn prediction models work better. They are using a dataset of customers who have left to build and test a model for predicting customer churn. This study explains a method that uses feature selection to find relevant features, which makes customer churn prediction models more accurate. The dataset with the results of feature selection is used to train and test the suggested model by using 10-fold cross validation for accuracy checks. To find out how well the Naïve Bayes method works for classification, you must first check how well the basic model works without any improvement.

C.Hybrid Classifier Methods

Hybrid classifier methods are created by combining two or more machine learning classifier algorithms. Hybrid classifiers have been developed in many types of field to enhance the prediction accuracy of models, as single predictor methods are not effective. Table 2 below shows the different hybrid machine learning classifier methods employed in predicting churn, highlighting their reliability and efficiency between 2003 and 2023.

Table 3. Hybrid Classifier

Hybrid Classifier	Definition
Voting Classifier	Combine the predictions from multiple classifiers and use a majority vote to make the final prediction. This can be implemented with techniques like hard or soft voting.
Stacking	Train multiple classifiers, and then use another model (meta-classifier) to combine their predictions. The meta-classifier learns how to best combine the outputs of the base classifiers.
Bagging and Boosting	Techniques like Bagging (e.g., Random Forest) and Boosting (e.g., AdaBoost, XGBoost) inherently combine multiple weak learners to create a strong learner. These methods can be effective in churn prediction tasks.
Feature-level Hybridization	Combine features extracted from different models to create a new set of features. This new feature set can be used as input to a single classifier. This approach is particularly useful when different models capture complementary aspects of the data.
Algorithm Switching	Develop a mechanism to switch between different classifiers based on certain conditions or thresholds. For example, one classifier may perform better on certain types of customers or under specific circumstances.
Sequential Hybrid Models	Train models sequentially, where the output of one model becomes the input or feature for the next model. This can be implemented in a pipeline fashion, allowing each model to refine the predictions made by the previous one.
Rule-based Systems	Combine machine learning models with rule-based systems. Rules can be derived from domain knowledge or heuristics and can be used to influence the final decision of the classifier.
Fusion of Outputs	Combine the output probabilities or scores from different classifiers using mathematical operations (e.g., averaging, weighted averaging). This is common in scenarios where the models provide probability estimates.
Neural Network Ensembles	Combine the outputs of multiple neural networks or models with different architectures to improve generalization and robustness

In 2007, Bong-Horng Chu and his research team developed a hybrid architecture that combines a learning mode with a usage mode. Using the Taiwan telecom dataset, the researchers applied the C5.0 algorithm or Decision Tree for classification and the Growing Hierarchical Self-Organizing Map (GHSOM) algorithm for clustering, resulting in an accuracy rate of 85% [10].

In 2009, Chih-Fong Tsai and colleagues developed a model that combined artificial neural networks (ANN) with both ANN and self-organizing maps (SOM) using hybrid algorithms. Using fuzzy testing data exclusively, this approach achieved accuracies of 94.32% and 93.06% without utilising feature selection approaches [5]. In 2010, Jiayin Qi et al. integrated ADTrees and Logistic Regression on a confidential telecom dataset, employing ROC as the assessment metric and highlighting the importance of variable and model selection in forecasting churn [11]. Bingquan Huang et al. (2010) [12] utilised a modified version of the NASA II approach to enhance the selection of sub-features on an actual Ireland Telecom dataset. They employed the Decision Tree algorithm to optimise the fitness function, which led to a significant 96% improvement in accuracy. In 2010, Wouter Verbeke and his team successfully combined AntMiner+ with ALBA [13], resulting in a specificity rate of 99.71% [12]. The most favourable outcomes were achieved when ALBA was merged with either RIPPER or C4.5. Lee et al. in 2010 [14] constructed a model utilising Partial Least Squares (PLS) approaches on the Tera Duke dataset, showcasing its higher performance in comparison to other individual classification models. Zhen-Yu Chen et al. came up with the HMK-SVM method in 2012 as a way to combine static and longitudinal trends in customer data [15]. When used on the Duke dataset, this method produced an amazing AUC value of 0.98. In 2013, Ying Huang and his team introduced a hybrid approach that merged the K-Means algorithm for customer segmentation with the FOIL algorithm for churn prediction. This combination yielded an AUC value of 89.70 [16]. Keramati et al. conducted a study in 2014 where they created churn prediction models using four algorithmswhich are Decision Trees (DT), Artificial Neural Networks (ANN), k-Nearest Neighbours (KNN), and Support Vector Machines (SVM) [17]. Among these techniques, ANN showed the best performance. The accuracy rate of 95% was achieved by integrating all the techniques stated earlier into a composite algorithm. Amjad Hudaib, Reham Dannoun and Osama Harfoushi introduced three hybrid models designed to assist service providers in analyzing and forecasting the future behavior of their clients in 2015 [3]. Three hybrid models are being examined to create precise and efficient churn prediction models, which will assist telecommunication firms in forecasting and analyzing the future actions of their consumers. There are two main parts to each of the three models which are the clustering phase and the forecast phase. Filtering of client info happens in the first stage. Predicting how customers will act is the next step. The first model looks at the Multilayer Perceptron Artificial Neural Networks (MLP-ANN) and the k-means method for filtering data. In 2016, Ammar A.Q et al [18]. proposed a hybrid firefly technique that outperformed traditional firefly algorithms by obtaining an accuracy of 86.3% in under 2.5 minutes. In their study, Wenjie et al. (2016) proposed the SDSCM hybrid method, which combines SCM and AFS. This approach achieved a clustering accuracy of 96% on both the Iris and wine datasets [5]. M Azeem et al. (2017) used fuzzy classifiers to look at the true positive rate and got an AUC value of 0.68 by using the OWANN classifier [19]. In their study, E. Sivasankar et al. (2017) conducted an analysis where they merged different clustering techniques. They discovered that the combination of the decision tree and K-Means led to improved accuracy [20]. In 2018, Adnan Amin and his peers came up with a way to use the distance factor of algorithms to get great results on four different sets of data. A method that depends on how far apart the models are. The method was tested on four different sets of data, and Naive Bayes was used as the standard predictor. The Bayesian Binomial method test was used to look at the whole system [21]. A study by J. Vijaya et al. in 2018 suggested a combined method called PPFCM with ANN for multi-class clustering. It had a very high success rate of 94% [22]. Arno De Caigny et al. (2018) proposed a hybrid technique called LLM for data classification, which achieved an AUC value of 0.62 [23]. Hoppner et al. (2018) created the Proftree predictor, which uses Decision Tree principles to make churn prediction models more accurate and easy to understand while also making them more profitable [24]. In their 2018 study, S. Babu et al [25]. proposed algorithms that specifically targeted the issue of class imbalance by enhancing the SMOTE and DT approaches. These algorithms achieved higher precision when applied to the UCI churn dataset.

D.Ensemble Classifier Methods

Ensemble classifiers are now essential tools in churn classification, utilizing the combined knowledge of numerous base classifiers to improve forecast accuracy. Random Forest is a highly popular ensemble method known for its resilience and effectiveness in handling high-dimensional data. Boosting and bagging are two well-known ensemble approaches that are highly acknowledged for their efficacy in churn classification problems. Boosting techniques, such as AdaBoost and Gradient Boosting Machines (GBM), train a series of weak learners in an iterative manner, with each learner specifically targeting the errors produced by the previous ones. Moreover, there have been the development of dedicated gradient boosting libraries like CatBoost to tackle the difficulties related to categorical features. CatBoost employs ordered boosting and oblivious trees to effectively manage category variables, hence improving both accuracy and speed in churn classification problems. Through the utilization of these ensemble classifiers, both researchers and practitioners may efficiently address the intricacies of churn prediction, resulting in valuable insights for client retention tactics. Table 3 below shows the different ensemble classifier methods employed in predicting churn, highlighting their reliability and efficiency between 2003 and 2023.

Table 4. Ensemble Classifier

Ensemble Classifier	Definition
Random Forest	Builds a lot of decision trees as an ensemble learning method. The final prediction is made by taking the average of the predictions from each tree or the forecast with the most votes. Random Forest is strong, works well with data that has a lot of dimensions, and doesn't overfit as often.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Gradient Boosting Machines (e.g., XGBoost, LightGBM)	Create a number of weak learners, which are usually decision trees, one after the other. Each tree fixes the mistakes made by the ones that came before it. Gradient boosting is often done with XGBoost and LightGBM, which are well-known for their fast speed and ability to handle data with complex relationships.
AdaBoost (Adaptive Boosting)	An ensemble that focuses on helping slow learners do better. It gives each data point a weight, and these weights are changed every time so that the wrongly classified cases stand out more. The end guess is the weighted sum of the guesses made by the weak learners.
Bagging (Bootstrap Aggregating)	When you bag, you train multiple copies of the same base classifier on different parts of the training data, which are usually selected with replacement. Most of the time, the end prediction is the average or majority vote of all the predictions. One kind of bagging algorithm is Random Forest.
Stacking	A meta-classifier is used in stacking to combine results from more than one base classifier. There is a plan to teach a meta-model that takes the results of the different classifiers as input and figures out how to best mix them. Stacking can find connections in the data that are more complicated.
Voting Classifier	The voting classifier is a simple ensemble method that works well. It takes the predictions of several classifiers and puts them all together. The final prediction is made by a majority vote (hard voting) or a weighted combination of the expected probabilities (soft voting).
Ensemble of Neural Networks	In deep learning, an ensemble of neural networks is a group of neural networks that are trained with different initializations or designs and then their predictions are put together. This can make generalization and stability better.

In 2006, Yong Seog Kim [26] suggested artificial neural networks (ANN) and logistic regression (Logit) methods should be used together to make feature selection prediction more accurate. Data from the Teradata Center for CRM at Duke University was used for this. Aurélie Lemmens and Christophe Croux [26] did a comparative review and found that Bagging and Boosting algorithms had strong predictive skills, particularly when dealing with huge datasets at 2006. A paper by Koen W. De Bock in 2012 called Rotboost and Rotation Forest ensemble models used methods like Principal Component Analysis (PCA) [27], Independent Component Analysis (ICA), and Supervised Principal Components Regression (SPR) to pull out features. Rotation forest and PCA were put together by De Bock to get an AUC value of 0.63. Koen W. De Bock et al. made a significant contribution in 2012 with the development of the GAMensplus algorithm. This algorithm achieved an accuracy rate of 63% on a European dataset. Additionally, the authors conducted a comparative analysis using Bagging, RSM, and Logistic Regression techniques [27]. In 2012, Adnan Idris and his colleagues investigated the combination of Genetic Algorithm with AdaBoost [28]. They used two commonly used datasets, namely the cell2cell and Tera datasets obtained from Duke University. By subjecting their strategy to a thorough assessment using a 10-fold cross-validation technique, they observed a significant AUC value of 0.89 [28], demonstrating the efficacy of their approach in predictive modeling. Concurrently, in their 2019 study, Adnan Idris and colleagues [29] suggested a method that combines Genetic Programming (GP) with Adaboost to achieve a more advanced degree of classification. Particle Swarm Optimization (PSO) was also used to fix the problem of classes that were not fair. during the same time frame. In their study, they looked into a different method that combined Random Forest, mRMR, and RF [1]. The alternate strategy produced an AUC value of 0.75, showing a significant but somewhat inferior performance compared to their amalgamation of Genetic Algorithm and AdaBoost. In addition, they performed an evaluation of reduced attribute sets using RF and KNN [28][1]. This helped us understand how different feature selection strategies might improve the performance of classifiers. In further research, Ning Lu et al [30]. did more study and used the Adaboost algorithm and Logistic Regression to make a model for predicting churn. Gradient Descent helped them get the best model, and the AUC number they got was 64.08. Recently, like in the work of J. Vijaya et al. (2018), researchers have been trying to improve churn prediction models by mixing rough set, wrapper, and filter techniques with ensemble methods like bagging, boosting, and random subspace. The objective is to improve feature selection [31]. Mohd Khalid Awang; Mokhairi Makhtar; Norlina Udin, et al [32]. utilized six distinct learning algorithms as the primary classifiers and evaluated various meta-model classifiers. The Multi-Layer Perceptron meta-model classifier was found to outperform the other classifiers. There is a significant amount of ongoing study in the field of classifier ensembles, with several papers suggesting different types of classifiers at both the base level and the meta-level, depending on the specific application being studied. This paper enhances the field of data mining research by proposing a potent amalgamation of base and meta-level classifiers for customer churn classification. In 2023 [33], Hoang Dang Tran, Ngoc Tuan Le, and Van-Ho Nguyen looked at the theories behind customer attrition and customer segmentation. They also suggested using supervised machine-learning methods to predict which customers will leave. According to the findings, the top two training methods are support vector machines and random forests. The cluster mean accuracy for random forest is 97% and the sample accuracy is 97.4%. Another finding is that client segmentation does not impact the capacity to foretell customer attrition. The results demonstrate that the random forest model performs admirably on the dataset, with an accuracy rate of approximately 97%. Following client segmentation, all of the models performed well in terms of mean accuracy; however, random forest achieved the best overall accuracy (97.25%), while logistic regression achieved the lowest (87.27%) [33].

E. Hybrid Ensemble Classifier Methods

Hybrid ensemble classifiers are put together in a new way to make hybrid ensemble classifiers. When you compare these classifiers to old-fashioned ways, they get the job done more accurately. Two or more ensemble methods are used together to make the system, such as boost-stacked and bagged-stacked. A hybrid ensemble is made by combining a lot of different classifiers with a lot of different ensemble methods. In Table 4, you can see the different hybrid ensemble classifier methods that were used to guess churn and how reliable and effective they were from 2003 to 2023.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 5. Hybrid Ensemble Classifier

Hybrid Ensemble Classifier	Definition
Ensemble of Ensembles	Combine different ensemble methods to create a higher-level ensemble. For example, you could combine the predictions of a Random Forest, AdaBoost, and a stacking ensemble to create a more robust and diverse model.
Ensemble with Feature Engineering	Integrate feature engineering techniques with ensemble methods. Extract features using various methods or from different sources, and then use an ensemble classifier to combine the predictive power of these features.
Ensemble with Feature Engineering	• Integrate feature engineering techniques with ensemble methods. Extract features using various methods or from different sources, and then use an ensemble classifier to combine the predictive power of these features.
Meta-Learning Ensemble	• Use meta-learning to build an ensemble of models that adapt to different subsets of the data. Each base model specializes in predicting certain patterns or segments of the customer base.
Sequential Hybrid Ensemble Models	Develop a hybrid ensemble by combining different models sequentially. For instance, use a Random Forest to pre-screen the data and then feed the predictions as additional features into a Gradient Boosting model.
Neural Network with Ensemble	Combine a neural network with ensemble methods. Train a neural network alongside an ensemble model, and use their predictions in combination to make the final decision. This can take advantage of both the non-linear representation learning capabilities of neural networks and the diversity offered by ensemble methods.
Rule-Based Hybrid Ensemble	Integrate rule-based systems or expert knowledge into the ensemble. Create rules based on domain knowledge or specific heuristics and combine them with the predictions from ensemble classifiers.
Ensemble with Algorithm Switching	Develop a mechanism to dynamically switch between different ensemble models or base classifiers based on certain conditions or performance metrics. This can be done during the training phase or at runtime.
Stacking with Neural Networks	Incorporate neural networks into a stacking ensemble. Train multiple base classifiers, including neural networks, and use a meta-classifier (e.g., logistic regression) to combine their predictions.

To predict customer attrition, Vijaya and Sivasankar employed particle swarm optimization (PSO) in conjunction with feature selection and simulated annealing (SA). With an F1 score of 96.06% and an accuracy of 94.08%, PSO-FSSA outperformed other popular machine learning algorithms when compared to their technique [31][22][24]. In the same year, Adnan and colleagues did a work that established a hybrid ensemble by combining diverse and similar classification techniques. Their research indicated that heterogeneous ensemble algorithms exhibit superior accuracy compared to individual and homogeneous ensemble approaches [21]. In 2018, Mahreen Ahmed and her coworkers used both baseline algorithms and a mix of ensemble methods, including boost stacked and bagged stacked techniques. Both datasets showed that the bagged stacked model worked very well, with accuracies of 98.4% and 97.2% [5]. During, 2018, Ammar and his team created ensemble stacking by combining benchmark techniques and implementing a cost-effective mechanism. The research was performed with the UCI churn dataset [25].

In 2023, Yan Tan, Ying Han Pang [35] purposes a heterogeneous ensemble model is suggested as a way to predict customer turnover in the industry field. Several trained base classifiers with different properties are put together using a stacking ensemble method. The suggested stacking ensemble method uses the unique strengths of each base classifier and the group's knowledge to make predictions more accurate by using a meta-learner. The results showed that the Chr-PmRF approach, which combines random forest, particle swarm optimization (PSO), and Minimum Redundancy and Maximum Relevance (mRMR), did a great job of predicting churn.

V. DATASETS FOR CHURN PREDICTION

Predicting churn has been used in both public and private datasets. Researchers get private churn datasets from different parts of the business world. But most of these private records can't be used because they are protected by intellectual property laws. Table 5 shows a list of publicly available datasets that have been used to predict customer churn, along with the number of study articles that have used these datasets.

Table 6. Datasets used for Customer Churn Classification

No	Datasets	Instances	Features
1.	UCI/Big ML – University of California	3333	21
2.	Kaggle-private dataset	100 000	100
3.	Cell2cell, Duke university Research Centre (CRM)	71,047	58

This article gives you a lot of ideas and ways to find clients who are about to switch service providers. A good churn prediction system should not only find likely customers who will leave, but it should also give a long-term estimate that is sufficiently detailed. When the marketing staff finds people who might leave, they usually get in touch with them. If it turns out that these people are likely to leave, the department does what it needs to do to keep them. The Boruta algorithm is used to find the most important factors in data analysis, which is used to measure service quality and find churn in the business world [28].

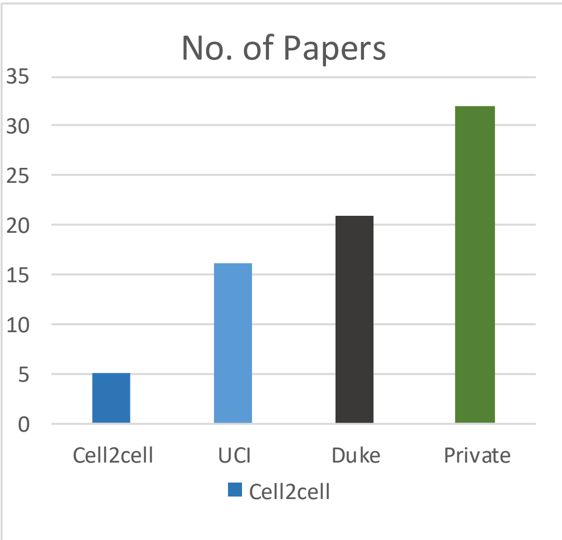

Figure 3. Datasets Versus Number of Papers

Retrieval Number: 100.1/ijrte.A803013010524
DOI: 10.35940/ijrte.A8030.13010524
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

In machine learning, feature selection, which is sometimes called "dimensionality reduction," is a popular method. In order to build a model, it includes picking out a subset of features from a dataset. The job is to pick the features that will help you solve the problem the best. Feature selection helps you understand data better by pointing out the most important traits and how they relate to each other. Picking out the most important factors in the data is part of it. The features that can be used to classify data are found and chosen with the help of an algorithm and a selection process. To make this possible, models are made simpler by cutting down on the number of variables and making them easier to understand. There are gaps in the user's writing [36][28][40][41].

VI. CHALLENGES AND FUTURE RESEARCH

The statistics that are available to the public are very uneven. Nature is what is being talked about. When it comes to predicting churn, the algorithms offered for this topic work very well. The class imbalance problem was addressed by Adnan et al [1]. through the use of linked classifiers and Particle Swarm Optimization (PSO). Bing Zhu and colleagues [37] addressed the issue of class imbalance in eleven separate datasets by employing the Random Under-Sampling (RUS) approach. When faced with two publicly available datasets that had unequal class distribution, Adnan et al [29]. utilized Particle Swarm Optimization (PSO) with an under-sampling strategy to resolve the issue. Another major issue is the process of creating a mixed classifier by mixing many classifiers. Since single predictors aren't very effective, mixed classifiers have replaced them. There are a number of proposed solutions to this issue. Third, you'll need to put together a hybrid ensemble using a variety of classifiers and ensemble techniques. The cutting-edge novel approach outperforms the more conventional hybrid classifier techniques. Choosing an appropriate feature to use as a churn prediction analytics guess is a major challenge.

VII. CONCLUSION

Research to keep valuable customers sometimes makes use of customer churn prediction, a popular area of study. There has been a recent influx of Machine Learning models applied to a wide range of public and private telecom information.

This article provides a thorough overview of several machine learning approaches that have been utilized between 2003 and 2023. Table 6 displays a collection of standard articles that were published from the year 2003 to 2023. Researchers, particularly in the telecom industry, have consistently developed churn prediction models, leading to an ongoing innovation in this subject. This study also discusses the existence of public and private churn statistics, as well as the significant issues faced by the telecom business. Hybrid ensembles are currently gaining popularity due to their superior predictive capabilities and significant importance. Figure 4 presents a comprehensive summary of the customer churn prediction studies conducted from 2000 to 2023.

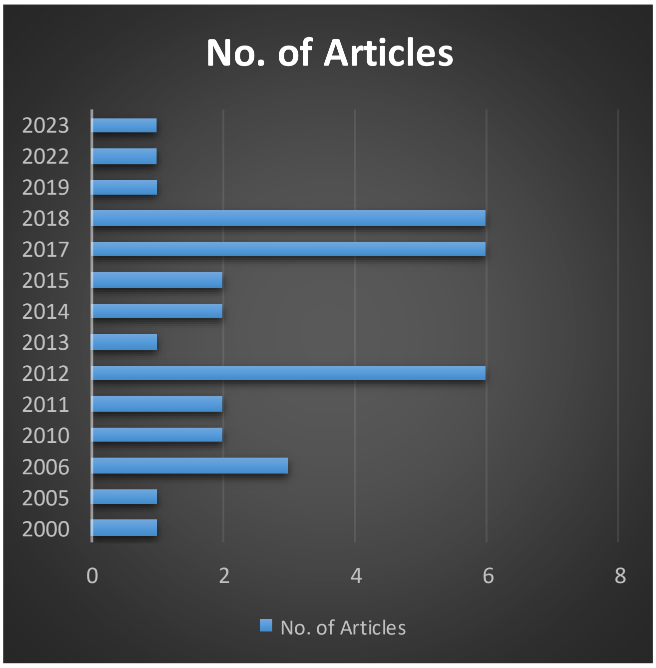

Figure 4. No of Published Articles Per Year in Customer Churn Prediction (2000 - 2023)

ACKNOWLEDGMENT

This work is supported by Fundamental Research Grant Scheme (FRGS/1/2023/ICT02/UNISZA/02/1) under the Ministry of Higher Education (MOHE) and University Sultan Zainal Abidin (UniSZA), Malaysia.

DECLARATION STATEMENT

Funding	Yes, This work is supported by Fundamental Research Grant Scheme (FRGS/1/2023/ICT02/UNISZA/02/1) under the Ministry of Higher Education (MOHE) and University Sultan Zainal Abidin (UniSZA), Malaysia.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors have equal participation in this article

REFERENCES

1. A. Idris, "Customer Churn Prediction for Telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," no. September, 2015, doi: 10.1109/INMIC.2012.6511498. https://doi.org/10.1109/INMIC.2012.6511498
2. X. I. A. Guo-en and J. I. N. Wei-dong, "Model of Customer Churn Prediction on Support Vector Machine," Syst. Eng. - Theory Pract., vol. 28, no. 1, pp. 71–77, 2008, doi: 10.1016/S1874-8651(09)60003-X. https://doi.org/10.1016/S1874-8651(09)60003-X
3. A. Hudaib, R. Dannoun, O. Harfoushi, R. Obiedat, and H. Faris, "Hybrid Data Mining Models for Predicting Customer Churn," no. May, pp. 91–96, 2015. https://doi.org/10.4236/ijcns.2015.85012

4. M. C. Mozer, R. Wolniewicz, and D. B. Grimes, "Churn Reduction in the Wireless Industry," no. January 1999, 2015.
5. J. Pamina, T. D. Rajkumar, S. Kiruthika, T. Suganya, F. Femila, and I. Introduction, "Exploring Hybrid and Ensemble Models for Customer Churn Prediction in Telecom Sector," vol. 3878, no. 2, pp. 299–308, 2019, doi: 10.35940/ijrte.A9170.078219. https://doi.org/10.35940/ijrte.A9170.078219
6. S. O. Abdulsalam, J. F. Ajao, B. F. Balogun, and M. Olaolu, "EAI Endorsed Transactions A Churn Prediction System for Telecommunication Company Using Random Forest and Convolution Neural Network Algorithms," vol. 7, no. 21, pp. 1–8, 2022. https://doi.org/10.4108/eetmca.v6i21.2181
7. G. Thakre, P. Wankhede, S. Patle, S. Joshi, and P. A. Chauhan, "Implementation of Machine Learning Model for Employee Retention Prediction," pp. 503–508.
8. I. Ullah, B. Raza, A. K. Malik, S. U. L. Islam, S. W. O. N. Kim, and M. Imran, "A Churn Prediction Model Using Random Forest : Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," IEEE Access, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999. https://doi.org/10.1109/ACCESS.2019.2914999
9. I. O. P. C. Series and M. Science, "Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes," 2020, doi: 10.1088/1757-899X/879/1/012090. https://doi.org/10.1088/1757-899X/879/1/012090
10. I. Pathan, N. A. Kanasro, F. B. Shaikh, M. U. R. Maree, and A. A. Chandio, "An Evolutionary Approach of Machine Learning for Monitoring Churn Prediction of Broadband Customer," vol. 10, no. 3, pp. 2623–2629, 2021. https://doi.org/10.30534/ijatcse/2021/1561032021
11. J. Qi et al., "ADTreesLogit model for customer churn prediction," pp. 247–265, 2009, doi: 10.1007/s10479-008-0400-8. https://doi.org/10.1007/s10479-008-0400-8
12. N. Hashmi, N. A. Butt, and M. Iqbal, "Customer Churn Prediction in Telecommunication A Decade Review and Classification," no. May 2014, 2013.
13. E. Radmehr and M. Bazmara, "A Survey on Business Intelligence Solutions in Banking Industry and Big Data Applications," vol. 7, no. 23, pp. 3280–3298, 2017.
14. H. Cho, Y. Lee, H. Lee, H. Lee, and Y. Lee, "Toward Optimal Churn Management : A Partial Least Square (PLS) Model Toward Optimal Churn Management : A Partial Least Square (PLS) Model," 2010.
15. Z. Chen, Z. Chen, Z. Fan, and M. Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data," Eur. J. Oper. Res., vol. 223, no. 2, pp. 461–472, 2012, doi: 10.1016/j.ejor.2012.06.040. https://doi.org/10.1016/j.ejor.2012.06.040
16. V. Jinde and P. A. Savyanavar, "Customer Churn Prediction System using Machine Learning," vol. 29, no. 5, pp. 7957–7964, 2020.
17. A. Keramati, R. Jafari-marandi, M. Aliannejadi, I. Ahmadian, and M. Mozaffari, "Improved churn prediction in telecommunication industry using data mining techniques," Appl. Soft Comput. J., vol. 24, pp. 994–1012, 2014, doi: 10.1016/j.asoc.2014.08.041. https://doi.org/10.1016/j.asoc.2014.08.041
18. M. Ewieda, E. M. Shaaban, and M. Roushdy, "Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry".
19. M. Azeem, M. Usman, and A. C. M. Fong, "A churn prediction model for prepaid customers in telecom using fuzzy classifiers," Telecommun. Syst., vol. 66, no. 4, pp. 603–614, 2017, doi: 10.1007/s11235-017-0310-7. https://doi.org/10.1007/s11235-017-0310-7
20. U. G. Inyang, O. O. Obot, M. E. Ekpenyong, and A. M. Bolanle, "Unsupervised Learning Framework for Customer Requisition and Behavioral Pattern Classification," vol. 11, no. 9, pp. 151–164, 2017, doi: 10.5539/mas.v11n9p151. https://doi.org/10.5539/mas.v11n9p151
21. A. Amin, B. Shah, A. Masood, F. Joaquim, and L. Moreira, "Just-in-time Customer Churn Prediction : With and Without Data Transformation International Journal of Information Management Cross-company customer churn prediction in telecommunication : A comparison of data transformation methods," Int. J. Inf. Manage., no. August 2020, pp. 0–1, 2018, doi: 10.1016/j.ijinfomgt.2018.08.015. https://doi.org/10.1016/j.ijinfomgt.2018.08.015
22. E. S. J. Vijaya, "Hybrid PPFCM-ANN model : an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network," Neural Comput. Appl., vol. 31, no. 11, pp. 7181–7200, 2019, doi: 10.1007/s00521-018-3548-4. https://doi.org/10.1007/s00521-018-3548-4
23. C. G. Mena, A. De Caigny, K. Coussement, K. W. De Bock, and S. Lessmann, "Churn Prediction with Sequential Data and Deep Neural Networks A Comparative Analysis ∗," pp. 1–12, 2019.
24. E. Stripling and B. Baesens, "Profit Driven Decision Trees for Churn Prediction," no. December 2017, 2018, doi: 10.1016/j.ejor.2018.11.072. https://doi.org/10.1016/j.ejor.2018.11.072
25. S. Babu and N. R. Ananthanarayanan, "Enhanced Prediction Model for Customer Churn in Telecommunication Using Enhanced Prediction Model for Customer Churn in Telecommunication Using EMOTE," no. February, 2018, doi: 10.1007/978-981-10-5520-1. https://doi.org/10.1007/978-981-10-5520-1
26. K. U. Leuven, "B AGGING AND B OOSTING C LASSIFICATION T REES TO Aurélie Lemmens and Christophe Croux".
27. K. W. De Bock and D. Van Den Poel, "Expert Systems with Applications Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models," vol. 39, pp. 6816–6826, 2012, doi: 10.1016/j.eswa.2012.01.014. https://doi.org/10.1016/j.eswa.2012.01.014
28. A. Idris and A. Khan, "Genetic Programming and Adaboosting based churn prediction for Telecom," no. October, 2012, doi: 10.1109/ICSMC.2012.6377917. https://doi.org/10.1109/ICSMC.2012.6377917
29. A. Idris and A. Iftikhar, "Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling," Cluster Comput., vol. 22, no. s3, pp. 7241–7255, 2019, doi: 10.1007/s10586-017-1154-3. https://doi.org/10.1007/s10586-017-1154-3
30. C. Science and Z. Zhang, "Using Combined Model Approach for Churn Prediction in Telecommunication Fa-Gui LIU, Zhi-Jie ZHANG*, Xin YANG," vol. 131, no. Eeeis, pp. 269–276, 2017.
31. J. V. E. Sivasankar, "An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing," Cluster Comput., vol. 22, no. s5, pp. 10757–10768, 2019, doi: 10.1007/s10586-017-1172-1. https://doi.org/10.1007/s10586-017-1172-1
32. M. K. Awang, M. Makhtar, N. Udin, and N. F. Mansor, "Improving Customer Churn Classification with Ensemble Stacking Method," vol. 12, no. 11, 2021. https://doi.org/10.14569/IJACSA.2021.0121132
33. H. Tran, N. Le, and V. Nguyen, "C USTOMER C HURN P REDICTION IN THE B ANKING S ECTOR U SING M ACHINE L EARNING -B ASED," vol. 18, pp. 87–105, 2023.
34. J. Vijaya and E. Sivasankar, "Computing efficient features using rough set theory combined with ensemble classification techniques to," Computing, vol. 100, no. 8, pp. 839–860, 2018, doi: 10.1007/s00607-018-0633-6. https://doi.org/10.1007/s00607-018-0633-6
35. T. Y. Lin et al., "Journal of Engineering Technology and Applied Physics Stacking Ensemble Approach for Churn Prediction : Integrating CNN and Machine Learning Models with CatBoost Meta-Learner," vol. 5, no. 2, pp. 99–107, 2023. https://doi.org/10.33093/jetap.2023.5.2.12
36. S. F. Sabbeh, "Machine-Learning Techniques for Customer Retention : A Comparative Study," vol. 9, no. 2, pp. 273–281, 2018. https://doi.org/10.14569/IJACSA.2018.090238
37. B. Zhu, B. Baesens, and K. L. M. Seppe, "An empirical comparison of techniques for the class imbalance problem in churn prediction," vol. 32, no. 0.
38. C. K. N, "RESEARCH ON CHURN PREDICTION IN MOBILE COMMERCE USING SUPERVISED MODEL .," no. 05, pp. 29–43, 2022, doi: 10.17605/OSF.IO/ZRX7H.
39. Sharma, N., Raj, A., Kesireddy, V., & Akunuri, P. (2021). Machine Learning Implementation in Electronic Commerce for Churn Prediction of End User. In International Journal of Soft Computing and Engineering (Vol. 10, Issue 5, pp. 20–25). https://doi.org/10.35940/ijsce.f3502.0510521
40. Thakur, T. B., & Mittal, A. K. (2020). Real Time IoT Application for Classification of Crop Diseases using Machine Learning in Cloud Environment. In International Journal of Innovative Science and Modern Engineering (Vol. 6, Issue 4, pp. 1–4).

https://doi.org/10.35940/ijisme.d1186.016420

41. Tamilarasi, Dr. A., Karthick, T. J., R. Dharani, & S. Jeevitha. (2023). Eye Disease Prediction Among Corporate Employees using Machine Learning Techniques. In International Journal of Emerging Science and Engineering (Vol. 11, Issue 10, pp. 1–5). https://doi.org/10.35940/ijese.c7895.09111023

AUTHORS PROFILE

Nurul Nadzirah Bt Adnan graduated with a Bachelor of Computer Science majoring in Security and Networking Computing from University Sultan Zainal Abidin, Terengganu, Malaysia. She also a research assistant and did her master degree at Faculty of Information and Computing in UniSZA.

Mohd Khalid Bin Awang graduated with a Bachelor of Science in Computer Science from Indiana University, Bloomington, USA. He further pursued a Master of Science in Information Technology and completed his PhD in Computer Science. His research areas encompass knowledge management, machine learning, ensemble methods, and artificial intelligence. With an active involvement in research and publishing, he has authored over 50 papers in peer-reviewed journals and holds positions as a reviewer for various conferences and journals.

Table 7. Summary of Customer Churn from 2000 to 2023 Survey on ML Based on Customer Churn Classification

Author/ Year	Application	Algorithm used	Application	Datasets
Michael et al. (2000)	Churn Classification	Logit Regression Neural Network	Private dataset 47,000 observations	ROC
Yu Zhao Bing Li (2005)	Churn Prediction	Support Vector Machine	Teradata Center for CRM at Duke University	Accuracy
Yong Seog Kim (2006)	Churn Classification	Ensemble of ANN and logit	Teradata Center for CRM at Duke University (100,000 examples)	Hypotheses and Coefficients
Aurelie et al. (2006)	Churn Prediction/ Classification	Bagging, stochastic gradient & binary logit	Teradata Center for CRM at Duke University	Top decile & Gini coefficien
Jiayin Qi (2010)	Churn Prediction/ Classification	Decision Tree and Logistic Regression	Private dataset	ROC
Marcin Owcza-Rczuk (2010)	Churn Prediction/ Classification	Logistic Regression	Private dataset (85,274 observations)	Lift curves
Bingquan Huang (2010)	Churn Prediction	Modified NSGA-II and C4.5	Ireland Telecom data (18,600 customers)	Overall Accuracy
Wouter Verbeke, David Martens (2010)	Churn Classification	ANTMINER+ AND ALBA	Public dataset (5000 observations)	Specificity
Adem Karahoca (2011)	Churn Classification	X-Means, Fuzzy C Means and Integrated With Anfis	Turkey GSM operator (24 900 GSM subscriber)	Sensitivity Specicify
Koen W.De Bock (2011)	Churn Prediction	Rotation forest and	European Telecom dataset (35,550 instances)	Accuracy
Adnan Idris (2012)	Churn Prediction	Genetic Algorithm with Adaboost	orange dataset (50,000 observations) and cell2cell dataset (40,000 samples)	AUC
Adnan Idris et al. (2012)	Churn Prediction	PSO+MRMR+RF	French telecom orange dataset	AUC
Koen W.De Bock et al. (2012)	Churn Prediction	GAMensplus	European dataset (35,550 observations)	AUC
Bingquan Huang et al. (2012)	Churn Prediction/ Classification	ANN, LR, DT, NB, SVM	Ireland telecom dataset (827,124 customers)	True & False churn rate
Wouter et al. (2012)	Churn Prediction	21 Classification Techniques	11 telecom datasets (both private & public)	AUC
Adnan Idris (2012)	Churn Prediction	ROTBOOST	Cell2cell (40000 instances)	AUC
Adnan Idris et al. (2013)	Churn Prediction	ROTBOOST+ + MRMR	Cell2cell (40000 instances) Tera Duke data (50,000)	AUC
YingHuang et al. (2013)	Churn Prediction	K-MEANS	Private dataset (104,199 customer records)	AUC
Ning Lu et al. (2014)	Churn Prediction	ADABOOST and Logistic Regression	(Private dataset)7190 customers	AUC
Keramati et al. (2014)	Churn Prediction	DT, ANN, KNN, SVM	Iranian mobile company. (3150 customer data)	Accuracy F-Score
Jin Xiao et al. (2015)	Churn Prediction	GMDH- NN	Churn (3333 observations)	Accuracy

Retrieval Number: 100.1/ijrte.A803013010524
DOI: 10.35940/ijrte.A8030.13010524
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Adnan Idris et al. (2015)	Churn Prediction	PSO, mRMR, Genetic Algorithm, Random Forest, Rotation Forest, RotBoost and SVM.	Orange datasets (50,000 observations) Cell2Cell (40,000 observations)	AUC
Adnan et al. (2017)	Churn Prediction	SVM, bagging, KNN, NB, NN	UCI dataset	Accuracy
M Azeem et al. (2017)	Churn Prediction	Fuzzy classifiers	south Asian Telecom (600000 Instances)	AUC
Adnan et al. (2017)	Churn Prediction	homo and heterogenous ensembles	UCI, KDD cup2009	AUC
Bing Zhu et al. (2017)	Churn Prediction/ Classification	RUS, SMOTE, Bagging RUS, SMOTE, Bagging,	11 data sets (4 public & 9 private)	AUC, EMP
E. Sivasankar et al. (2017)	Churn Prediction	FCM, PFCM & K-Means, DT	Churn dataset (50,000 observations)	Accuracy
E. Sivasankar et al. (2017)	Churn Prediction	PSO, NB, SVM, Random Forest and other hybrid models	Orange Small and Orange Large	Accuracy
J. Vijaya et al. (2018)	Churn Prediction	Baseline classifiers, Bagging, Boosting, RS, rough set, filter and wrapper	Teradata Centre for CRM at Duke University	Accuracy
J. Vijaya et al. (2018)	Churn Prediction	PPFCM-ANN	Duke Tera Data	Accuracy
J. Vijaya et al. (2018)	Churn Prediction	Fuzzy clustering algorithms with baseline classifiers	Private dataset	Accuracy
Adnan Amin et al. (2018)	Churn Prediction/ Classification	CCP method with distance factor	UCI Churn (3333 Observations), IBM Watson (7043 observations), Abinav Kaggle (100,000 records) and Pakdd2006(18,000 records)	Accuracy and F-Measure
Ammar et al. (2018)	Churn Prediction	Ensemble stacking	UCI Churn dataset	Accuracy
Mahreen Ahmed et al. (2018)	Churn Prediction	Boosted-Stacked Bagged-Stacked	Boosted-Stacked Bagged-Stacked	Accuracy
Chitra Kiran. N (2022)[38]	Churn Prediction	ML model, Feature selection Heatmap Model SVM, NB LDA	Data Collection Kaggle Eliminate null Statistics values	Accuracy
Praveen Lalwani (2022)	Churn Prediction/ Classification	Logistic Regression, Naïve Bayes, Support Vector Machines, Decision Trees, Random Forest, XGBoost Classifier, CatBoost Classifier, AdaBoost Classifier and Extra tree Classifier	Private dataset	AUC
Hoang Tran (2023)	Churn Prediction	K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, And Support Vector Machine	Kaggle Datasets	Accuracy