Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 24s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Review On Machine Learning Methods For Over-The-Top (OTT) Platforms Customer Churn Prediction

Bathula Prasanna Kumar[1, 2], Dr. Edara Sreenivasa Reddy[3]

[1]Research Scholar, Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

Email ID: prasannabpk@gmail.com

[2]Associate Professor, Department of Computer Science and Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India.

[3]Professor, Department of Computer Science and Engineering, VIT-AP, India.

Email ID: sreenivasareddy.e@vitap.ac.in

Cite this paper as: Bathula Prasanna Kumar, Dr. Edara Sreenivasa Reddy, (2025) A Review On Machine Learning Methods For Over-The-Top (OTT) Platforms Customer Churn Prediction. *Journal of Neonatal Surgery*, 14 (24s), 111-123.

ABSTRACT

Market deregulation and globalization have significantly heightened competition across industries, resulting in evolving market dynamics and an increase in customer churn. For OTT service providers, accurately predicting and addressing churn is crucial for retaining subscribers and ensuring sustainable growth. This study examines 185 published articles from 2018 to 2024, focusing on customer churn prediction through machine learning techniques. In contrast to previous reviews that often explore isolated aspects of churn prediction, this research places particular emphasis on OTT-related sectors, including telecommunications, finance, and online streaming platforms. It highlights the importance of creating new datasets and developing predictive models tailored to these industries. The findings reveal a notable research gap regarding the profitability aspect of churn prediction models. To address this, the study advocates for the integration of profit-based evaluation metrics, enabling better decision-making, improved subscriber retention strategies, and enhanced profitability. This comprehensive approach underscores the importance of aligning predictive techniques with actionable business outcomes in the OTT churn landscape.

Keywords: *Churn Prediction, Machine Learning, Decision Making, Customer Defection, Marketing Analytics*

1. INTRODUCTION

In recent years, Over-the-Top (OTT) media services have revolutionized the media landscape. They enable viewers to stream content anytime, anywhere, bypassing traditional cable, broadcast, and satellite TV platforms[1]. This shift has also made it easier and more cost-effective to target specific audiences with marketing efforts. OTT platforms provide an affordable way to access content, often eliminating the need for costly contracts or monthly fees. Users typically sign up for an account through a website or app and gain access to a wide range of content, from live TV channels to on-demand shows and movies[2]. Additionally, OTT services often offer features like saving favorite shows, receiving personalized recommendations, and accessing exclusive content.

OTT consumption has significantly grown in India, driven by factors such as income, age, ad-free viewing, and convenience[3]. OTT platforms operate on three models: Advertising Video on Demand (AVOD), Subscription Video on Demand (SVOD), and the Freemium Model, which combines both. Subscription services, in which users pay for recurring access, have become increasingly popular and adaptable across industries[4]. A major challenge for OTT services is subscriber churn, which refers to the rate at which users cancel their subscriptions. High churn rates reflect customer dissatisfaction and impact retention. Causes of churn range from rising acquisition costs to lack of personalized content and pricing issues. To tackle churn, companies must deeply understand their customer base and the reasons behind attrition. This is where machine learning plays a critical role. By analyzing large amounts of subscriber data, machine learning technologies help broadcasters identify patterns in consumer behavior, preferences, and trends that are likely to lead to churn[5].

Figure.1. OTT Customer Churn

A. Need of Churn Prediction in OTT Business

In the highly competitive OTT market, economic success hinges on extending the average customer's lifespan and increasing consumption. Churn prediction models play a pivotal role in identifying at-risk customers, enabling companies to implement targeted retention strategies that safeguard their customer base and optimize revenue[7].

Retention of subscribers minimizes the need for OTT platforms to invest heavily in advertising for new users, enabling them to focus on nurturing existing relationships[8].

- Long-term subscribers who remain engaged with the platform and are satisfied with the service are more likely to consume additional content and recommend the platform to others.
- Leveraging the data collected over time, engaging and retaining loyal subscribers becomes more cost-effective.
- Subscribers who stay with the platform for an extended period typically contribute more value and revenue.
- Conversely, subscriber churn represents a significant cost for OTT platforms, reducing the profitability of acquiring new users to compensate for the losses.

B. Why is Machine Learning is used for OTT Customers Churn Prediction

Machine Learning (ML), a vital computer science discipline, helps uncover hidden patterns within large datasets by leveraging statistical, mathematical, artificial intelligence, and machine learning techniques. Despite organizations having access to vast data reserves, the primary challenge resides in distilling this data into actionable insights, a task that data mining tools are well-equipped to tackle. OTT churn prediction is fundamentally a classification problem that distinguishes between regular subscribers and those likely to churn[8].The advantages of utilizing Machine Learning (ML) within the field of OTT Customers Churn Prediction can be succinctly outlined as follows:

- ML has provided a faster way to analyze large amounts of data.
- ML significantly reduces future search space, making it a powerful and efficient tool for OTT churn prediction.
- ML advancements have led to innovative methods that improve the accuracy of churn detection, helping to identify at-risk subscribers early.

C. Motivation

While substantial interest in OTT churn prediction has led to numerous review articles on the topic, a notable gap remains in providing comprehensive guidelines for businesses seeking to develop effective OTT churn prediction models. This knowledge gap highlights the need for an in-depth survey offering insights into the end-to-end pipeline for predicting customer churn in OTT platforms. This necessity forms the foundation of our comprehensive survey, guided by a primary research question driving our exploration[10].

- "How can businesses, particularly those in the OTT industry, leverage insights and methodologies from current Machine Learning research to effectively predict customer churn?"

To clarify the motivation behind our survey and research question, this study summarizes the focus areas of review articles published between January 2020 and January 2024, providing a comparative analysis with our work. It identifies critical gaps in the OTT churn prediction landscape, particularly in the application of Machine Learning models, the integration of profit-based metrics for model validation, and the contextualization of trends in predictive modeling. This survey aims to bridge these gaps and deliver a comprehensive understanding of the OTT churn prediction ecosystem[11].

The methodology used for this literature review is outlined in Section II. Section III provides a detailed overview of Machine Learning techniques. Section IV examines the general evaluation methods for these techniques. Section V highlights validation strategies for predictive models. Finally, Section VI synthesizes the identified gaps and presents recommendations, contributing to the existing body of knowledge[12].

2. METHODOLOGY

A three-step systematic literature review was conducted to provide an unbiased and objective evaluation of the current state-of-the-art in OTT customer churn prediction and the future applications of machine learning in this domain. First, the scope

of the study was refined to focus exclusively on articles leveraging machine learning for predicting OTT customer churn. Next, relevant articles were identified and selected from established databases by employing targeted search terms and iterative query refinement. Finally, the results and key insights from the selected studies were analyzed and synthesized to derive meaningful conclusions[13].

A. Articles Collection

To ensure comprehensive and diverse topic coverage, an initial literature scan was conducted using the primary terms *"OTT Customer Prediction"* and *"Machine Learning"* across databases such as Web of Science (WoS), DBLP, Scopus, and Google Scholar to identify relevant keywords. The search terms were then expanded by including additional keywords, such as *"OTT Customer defection," "OTT Customer turnover," "OTT Customer switching,"* and *"OTT Customer abandonment."* These expanded terms were subsequently combined with *"Machine Learning"* or *"Artificial Intelligence"* to refine and optimize the final search query[14]. Next, a thorough analysis of the articles' abstracts, keywords, and key contributions was performed. Only those articles that met the inclusion criteria. This study adhered to the PRISMA framework (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) as outlined in [43] to identify relevant research articles. The PRISMA process consists of four key stages: identification, screening, eligibility, and inclusion. These steps are visually represented in Figure 2 through a flow diagram. Initially, a total of 681 articles were retrieved from bibliographic databases using specific keywords, along with an additional 35 articles sourced from other platforms. After removing duplicates and excluding ineligible studies for various reasons, 479 articles remained. Of these, 185 articles were screened, while 294 were excluded due to unavailability. Ultimately, 185 articles were included in the review for this study[15].

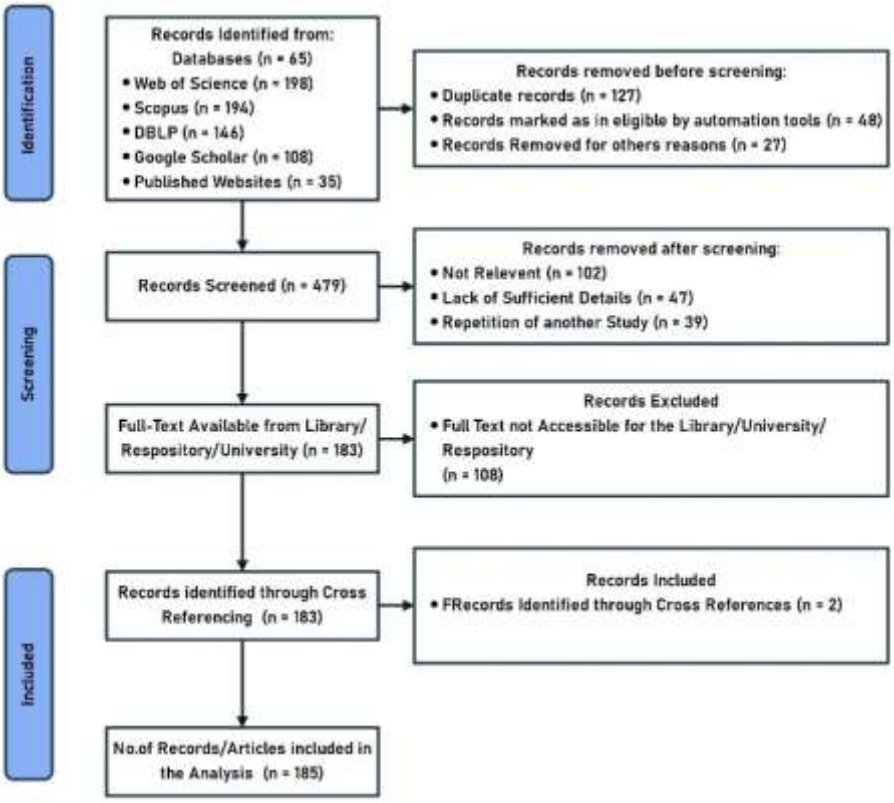

Figure.2. Flowchart illustrating article count and inclusion/exclusion criteria in the study.

Next, a thorough analysis of the articles' abstracts, keywords, and key contributions was performed. Only those articles that met the inclusion criteria as

- Articles relevant to churn prediction using Al or ML techniques
- Articles published between 2018 and Jan-2024
- Relevant articles providing sufficient details and not duplicating other studies

Ultimately, this literature review condensed to a count of 185 articles, and Figure.3 and 4 show distributional statistics concerning the total number of articles.

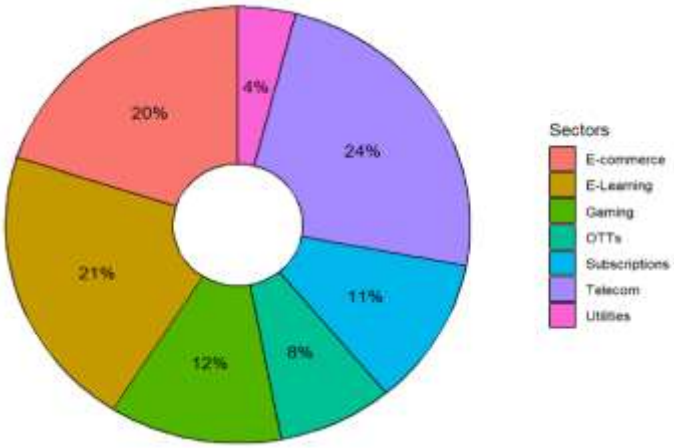

Figure.3. Sector wise distribution of articles

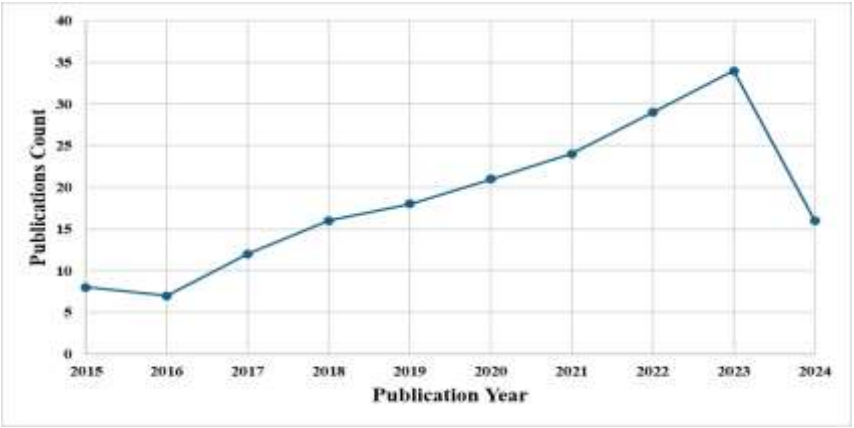

Figure.4. Yearly Publications trend (Published until Nov-2024).

3. MACHINE LEARNING ALGORITHMS

Machine Learning is the motivation of computers to learn from data to perform certain tasks as humans do. Machine Learning algorithms use sample data, known as "training data," to build prediction models without being programmed. Based on the model learning method, Machine Learning approaches are divided into different categories like Supervised Learning, Semi-Supervised Learning, Unsupervised Learning, Reinforcement Learning, Ensemble Learning, Transfer Learning and Instance Based Learning and Deep Learning[16].

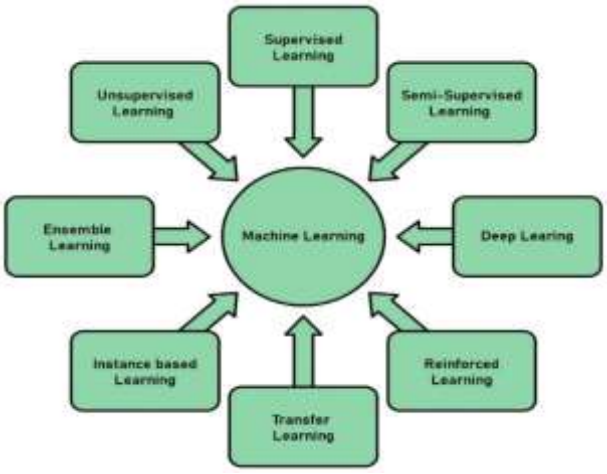

Figure.5. Different Machine Learning Techniques

A. Supervised Learning

Supervised machine learning algorithms are those algorithms which needs external assistance. The input dataset is divided into train and test dataset. The train dataset has output variable which needs to be predicted or classified. All algorithms learn some kind of patterns from the training dataset and apply them to the test dataset for prediction or classification[17]. The main tasks of supervised learning are Classification and Regression:

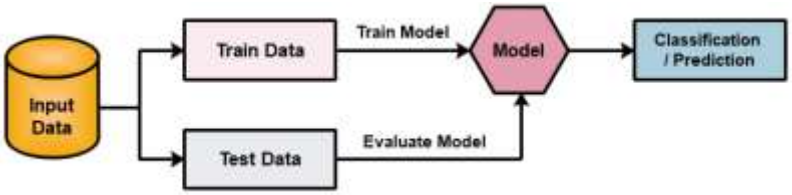

Figure.6. Supervised Learning Workflow

B. Unsupervised Learning

In unsupervised learning, the data is unlabeled, so the learning algorithm is left to find commonalities among its input data. The goal of unsupervised learning is to discover hidden patterns within dataset, which allows the computational machine to automatically discover the representations that are needed to classify raw data.

Since there are several such factors involved in predicting OTT Customer Churn Prediction, many researchers are aiding the help of machine learning techniques.

- **Logistic Regression:** This algorithm estimates discrete values (0 or 1) based on a given set of an independent variable(s). It basically predicts the probability of occurrence of an event by fitting data to a logit function. Logistic Regression uses the logistic function f(z) = 1 / (1 + e^(-z)) (where z represents the linear combination of the feature values and their respective weights) and the logistic loss function.
- **Support Vector Machine:** In SVM, each data item is plotted as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.
- **Naïve Bayes:** Assumes that the features are conditionally independent of each other, learns the class probabilities (P(y)), and calculates the likelihood probabilities (P(X|y)) based on the observed features in the training data.
- **Decision Tree:** The algorithm continues splitting the data at each node, creating child nodes and branches, until a termination condition is met. Gini Impurity and Information are the popular criteria for splitting data.
- **Random Forest:** Each decision tree is trained independently using the random subsets of data and features, and the combined forecasts of all the decision trees are used to arrive at the final prediction.
- **Bagging Classifier:** Based on a distinct bootstrapped subset of the training data, each base classifier is independently trained. The sum of the forecasts from each model is the final prediction.
- **AdaBoost Classifier:** A weighted voting system is used to decide the final prediction after each weak classifier is trained on a subset of the training data and given a weight depending on classification accuracy.
- **Gradient Boosting Classifier:** Follows a boosting framework where weak learners are trained sequentially to correct the errors of previous models by utilizing gradient descent optimization.
- **XGBoost Classifier:** Follows the gradient boosting framework and uses decision trees as weak learners, which are constructed in a greedy manner.
- **Multi-Layer Perceptron:** Uses a process called backpropagation to optimize the weights of the connections between neurons iteratively. Uses optimization algorithms like gradient descent.

4. MODEL VALIDATION

After developing a Machine Learning model, the next critical step is evaluating its performance. Model validation involves assessing the model's ability to make accurate predictions on unseen data. A fundamental principle of model validation is to partition the dataset, reserving a portion exclusively for validation. This reserved data, not utilized during training, is used to measure the model's predictive performance. The effectiveness of the model can be quantified using various metrics, such as accuracy, precision, and recall. In this section, we delve into different concepts and techniques associated with model validation[50].

A. Cross-Validation

Cross-validation (CV) is a commonly used sampling method to evaluate the generalizability of a model over a dataset. Its core principle is to use all available data to train and test the model.

- ***K-fold Cross-Validation:*** This technique involves randomly dividing a dataset into *K* subsets of equal size. Each subset is used once as a validation set, while the remaining subsets are combined to train the model. A special case of this method is Leave-One-Out Cross-Validation (LOOCV), where the model is trained and evaluated on each individual instance in the dataset. While LOOCV is highly exhaustive, it comes with significant computational complexity. *K*-fold cross-validation, on the other hand, is recognized for its lower variance compared to other methods, though it may exhibit slightly higher bias. This approach helps mitigate overfitting by partitioning the data into distinct subsets. However, despite its strengths, *K*-fold CV has certain limitations. Its primary drawback lies in the computational demands of repeatedly training and testing the model. Furthermore, if the data distribution across the folds is imbalanced, it can introduce bias into the evaluation process, potentially compromising the validity of the results[51].

B. Test Dataset

It's a common practice to split the data into 80% training and 20% test data. In data science, a typical split is two-thirds of the data for training and one-third for test. Numerous studies have effectively utilized test datasets to obtain out-of-sample performance estimates and validate the accuracy and reliability of their churn prediction models[48].

Performance Metrics

Algorithms discussed previously are evaluated using the following metrics including False Negative (FN), False Negative (FN), and False Positive (FP) false positives (FP).

- **Sensitivity:** It measures proportions among actual positives that are accurately spotted. TP represents True Positive, whereas FN represents False Negatives.

$$\textbf{Sensitivity(SE)}=\frac{\textbf{TP}}{\textbf{TP+FN}}$$

- **Specificity:** Classifying patients who do not have the disease is the most important part of the diagnostic process.

$$\textbf{Specificity(SP)}=\frac{\textbf{TN}}{\textbf{FP+TN}}$$

True Positive (TP) and False Negative (FN) are used in this context[55].

- **Accuracy:** It is the result of a mix of systematic and random errors. Accuracy's high value necessitates high precision values. For evaluated images, it calculates the proportion of true positive and true negative[55].

$$\textbf{Accuracy (Acc):}\frac{\textbf{TP+TN}}{\textbf{TP+FN+TN+FP}}$$

- **The Receiver Operating Characteristics and The Area Under its Curve**: The receiver operating characteristics (ROC) and the area under its curve (AUC or AUROC) are commonly used evaluation metrics in the churn prediction literature. These metrics provide an objective and systematic way to assess classifier performance without requiring a specific classification threshold value, as multiple values are tested. It is a function of sensitivity to the inverse of specificity[56].

$$\textbf{ROC}=\frac{\textbf{Sensitivity}}{\textbf{1}-\textbf{Specificity}}$$

A ROC curve is a graphical representation that plots sensitivity (true positive rate) on the y-axis against the inverse of specificity (false positive rate) on the x-axis. In the context of OTT churn prediction, a value of 1 indicates that the model can effectively differentiate between churn and non-churn customers, while deviations from this value reflect a decline in performance. Unlike accuracy, the ROC curve is better suited for imbalanced datasets, as the AUC (Area Under the Curve) evaluates the classifier's ability to distinguish between churn and non-churn customers based on predicted probabilities. The AUC provides a comprehensive summary of a classifier's performance across all possible decision thresholds, making it a popular metric in OTT churn prediction. However, in cases of highly imbalanced datasets—common in churn scenarios—the ROC curve may be biased toward the majority class. As a result, relying solely on ROC/AUC can lead to misleading conclusions. It is, therefore, essential to use additional evaluation metrics alongside ROC/AUC to ensure the robustness and effectiveness of OTT churn prediction models[56].

- **Lift:** is an essential metric in OTT churn analysis, used to categorize customers based on their likelihood of churning. It helps prioritize customers with a higher probability of churn by placing them in the top segment. The lift value of a target group indicates the proportion of customers who actually churn within that group compared to the overall customer base. For example, if 2% of the overall customer population discontinues the service, but within the identified "churners" group, 8% drop the service, the lift is calculated as 4. This demonstrates the model's effectiveness in identifying high-risk customers, making lift a valuable tool for targeted retention strategies in the OTT domain. Let S be a ranked list of customers based on churn score, then the lift index is calculated as:

$$\text{Lift Index} = \frac{1.0 \times S_1 + 0.9 \times S_2 + ... + 0.1 \times S_{10}}{\sum_{i=1}^{10} S_i}$$

where S_i is the number of churners in the i^{th} decile of the list S, and its range is between 0.5 and 1. An optimal lift index is 1 with $S_1 = \sum_{i=1}^{10} S_i$. S_i if the churn rate is lower than 10%, whereas 0.5 indicates the selection of a random customer as a churner. While ROC curve analysis and lift-based methods are widely used in the churn literature, they do not consider the cost and expected profit of a retention campaign. Therefore, it is important to consider other factors when making decisions about marketing strategies and customer retention[47].

5. LITERATURE SURVEY

M.A.H. Farquad [4] proposed a hybrid approach to address the limitations of standard SVM models, which function as black boxes and fail to present the knowledge gained during training in an interpretable form. The approach involves three phases: Feature Reduction: SVM-Recursive Feature Elimination (SVM-RFE) reduces the feature set. Model Development: The reduced dataset is used to build the SVM model, and support vectors are extracted. Rule Generation: Rules are derived using a Naive Bayes Tree (NBTree), which combines decision trees with a Naive Bayesian classifier. The approach was tested on the highly imbalanced Business Intelligence Cup 2004 dataset, representing bank credit card customers with 93.24% loyal and 6.76% churned customers. While effective, the model lacked scalability for large datasets.

Lee et al. [11] developed an accurate and concise predictive model for churn prediction using a Partial Least Squares (PLS) approach on datasets with high variable correlation. Their methodology not only predicts customer churn behavior effectively but also integrates a straightforward and practical churn marketing program. This approach enables marketing managers to maintain an optimal—or near-optimal—level of churners efficiently through targeted marketing efforts, with PLS serving as the core prediction modeling technique.

Koen W. De Bock [12] proposed GAMensPlus, an ensemble classifier based on generalized additive models (GAMs), designed to balance performance and interpretability for churn prediction. GAMens, built using Bagging, the Random Subspace Method, and semi-parametric GAMs as base classifiers, was enhanced with tools for interpretability: generalized feature importance scores and bootstrap confidence bands for smoothing splines. Experiments on datasets from six real-world churn prediction projects demonstrated that GAMensPlus delivers strong classification performance, matching or exceeding the performance of individual classifiers like logistic regression and GAM.

Kriti [13] in her paper Customer churn: A study of factors affecting customer churn using machine learning has been too successful to find out various factors affecting customer churn (price sensitivity, technology, customer service, tenure, security). Also comparing various algorithms to analyze customer churn and prescribe solutions to avoid this churn. She has given the future work as the predictions from the ML model can help in understanding the customers who might leave their services. Also suggested various solutions based on those predictions.

Essam Abou El Kassem and Shereen Ali Hussein [14] in their paper has explained that Customer churn is a problem for most companies because it affects the revenues of the company when a customer switches from a service provider company to another. They've used social media sentiment analysis to predict the factors behind customer churn.

Praveen Lalwani and Manas Kumar Mishra [15] in their paper has Compared the time taken to train the model and accuracy of various ML algorithms. Their paper concludes that ensemble learning techniques such as XGBoost classifier gives maximum accuracy when compared to other models.

Saran Kumar A. [16] In his paper had conducted a survey on various ML algorithms and techniques to predict customer attrition or churn rate. Proposed to use various boosting classification techniques for better accuracy.

Pradeep B and Sushmitha Vishwanath Rao [17] in their paper have explained how to use various ML algorithms to analyze customer attrition or churn rate in the logistics industry.

Amiya, Ranjan et al. [24] employed various machine learning models, including Logistic Regression, Decision Trees, Random Forest, and XGBoost Classifier, for customer churn prediction. Among these, the XGBoost Classifier achieved the

highest accuracy of 89%, outperforming the other models tested in the study. The paper primarily focuses on the application of various machine learning models to predict customer churn rates but does not explore the potential impact of external factors such as market trends, competitive landscape, or economic conditions on customer retention, which could provide a more comprehensive understanding of churn dynamics.

Tasneem et al.[25] utilized various machine learning algorithms, including Decision Trees, Random Forests, Logistic Regression, Support Vector Machines, Naïve Bayes, and Neural Networks, for customer churn prediction. The Decision Tree algorithm outperformed the others, achieving an accuracy of 96.7%, precision of 96.9%, recall of 99.3%, and an F1-score of 98.1%, highlighting its effectiveness in churn prediction. Further investigation could be conducted to explore the impact of different hyperparameters on the performance of the predictive models, as well as the comparison of additional machine learning algorithms beyond the ones mentioned in the paper to identify potential improvements in churn prediction accuracy.

Victor Chang et al.[26] analyzed and forecasted customer churn in the telecommunications industry using Decision Trees, Boosted Trees, and Random Forests. Among these, the Random Forest model achieved the best performance, with a predictive accuracy of 91.66%, precision of 82.2%, and recall of 81.8%, demonstrating its effectiveness in churn prediction for this sector. While the study highlights the effectiveness of ensemble learning models like Random Forests in predicting customer churn behavior, it does not discuss potential challenges or drawbacks associated with implementing these models in real-world business settings.

Muteb et al. [27] employed various machine learning models, including Random Forest, LightGBM (LGBM), XGBoost, Logistic Regression, Decision Trees, and a custom Artificial Neural Network (ANN) model, to predict customer churn in the telecommunications sector. The ensemble averaging method achieved an accuracy of 0.79 and a recall of 0.72 on the test data, slightly underperforming compared to the standalone LGBM, XGBoost, and Logistic Regression models. The study mentions that the ensemble averaging method achieved a slightly lower accuracy and recall in the test data compared to LGBM, XGBoost, and Logistic Regression, indicating a limitation in the performance of the ensemble method.

Opara John et al.[28] conducted a comparative evaluation of AdaBoost, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) for customer churn prediction in the telecommunications industry. Among these, XGBoost demonstrated superior performance, achieving the highest accuracy of 89.51% and a recall rate of 92.48%, outperforming both Gradient Boosting and AdaBoost. The study highlights the challenge of imbalanced data in customer churn prediction, which can lead to biased model performance. To address this limitation, the Synthetic Minority Oversampling Technique (SMOTE) is employed as a strategy to balance the dataset and improve the accuracy of the predictive models.

Xiayu Li et al. [29] utilized a stacking ensemble method incorporating CatBoost for customer churn prediction in the telecommunications industry. The CatBoost model outperformed other machine learning models, achieving the highest accuracy of 0.8119, demonstrating its effectiveness in this domain. The paper mentions the need for continuous monitoring and iteration in churn analysis to adapt to new data or changes in business conditions, indicating a limitation in the static nature of the predictive model. This implies that the model may not fully capture dynamic shifts in customer behavior over time.

Muhammad Maulana et al.[30] applied various machine learning models, including Logistic Regression, Random Forest, Support Vector Machine (SVM), Gradient Boosting, AdaBoost, and Extreme Gradient Boosting (XGBoost), for customer churn classification. The XGBoost algorithm outperformed the others, achieving the highest accuracy of 0.829424 in identifying churn customers. While the paper addresses the challenge of class imbalance using the SMOTE method, it does not investigate alternative techniques or strategies for handling class imbalance, such as different sampling methods or cost-sensitive learning approaches, which could provide insights into improving model performance and accuracy.

Manal et al.[31] employed several predictive models, including Random Forest, k-Nearest Neighbor (KNN), and Support Vector Machine (SVM), for churn prediction in the telecommunications industry. These models demonstrated high accuracy, with results surpassing 79% on the test set. The highest AUC score of 84% was achieved by the SVM and Bagging classifiers when used in an ensemble method, outperforming the other models. The paper highlights the difficulty in predicting customer churn due to the great influence of individuals and products, which makes it challenging to establish clear rules for prediction.

Liwei Chen et al.[32] applied the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance in the training set, enhancing the accuracy of their prediction model. The Random Forest (RF) method outperformed the Support Vector Machine (SVM) in predicting customer churn, achieving an accuracy of 0.80, an F1 score of 0.76, and an Area Under Curve (AUC) of 0.93, demonstrating its effectiveness in analyzing customer turnover data. Lack of discussion on the interpretability of the machine learning models used, as understanding the reasons behind the predictions made by the models is crucial for telecom companies to take appropriate actions to reduce customer churn.

Naveen Kumar R et al. [33] evaluated four classification algorithms—Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), and a Hybrid Algorithm (HA) that combines DT and RF—for predicting customer churn in the telecom industry. The study found that the Hybrid Algorithm (HA) achieved the highest accuracy of 95%, outperforming the

individual models: 85% for Logistic Regression (LR), 91% for Decision Tree (DT), and 94% for Random Forest (RF). The paper focuses primarily on the effectiveness and performance of the classification models in predicting customer churn in the telecom industry, without delving into potential drawbacks or challenges faced during the analysis.

Angela Yi Wen et al. [34] utilized K-Nearest Neighbors (KNN), Random Forest (RF), AdaBoost, Logistic Regression (LR), XGBoost, and Support Vector Machine (SVM) for customer churn prediction. The study found that XGBoost was the most effective classifier, achieving an accuracy of 79.67%, precision of 64.67%, recall of 51.87%, and an F1-score of 57.57%. The paper does not explicitly mention any limitations regarding the predictive models, or the dataset used.

Abdelrahim et al. [35] used Decision Tree, Random Forest, Gradient Boosted Machine Tree (GBM), and Extreme Gradient Boosting (XGBoost) for churn prediction. The model achieved an Area Under Curve (AUC) value of 93.3%, indicating high accuracy in predicting customers likely to churn. This performance was significantly improved by incorporating Social Network Analysis (SNA) features, which boosted the AUC from 84% to 93.3%. The dataset encountered challenges such as a significant number of variables being nearly constant or having very few unique values. Specifically, about 50 numeric variables contained one or two discrete values, and nearly 80 categorical variables had less than 10 categories. Additionally, 77 of the numerical variables had more than 97% of their values filled with 0 or null, indicating that many variables could be removed due to their lack of variability.

Alae Chouiekh et al.[36] used Deep Convolutional Neural Networks (DCNN) for customer churn prediction. The study found that DCNN achieved an F1 score of 91%, outperforming traditional machine learning algorithms such as Support Vector Machines, Random Forest, and Gradient Boosting Classifier when predicting churn from a dataset of 18,000 prepaid subscribers. Deep Convolutional Neural Networks (DCNN) to predict customer churn, there is no discussion on the potential integration of other advanced machine learning techniques or hybrid models that could leverage the strengths of both DCNN and traditional algorithms to further improve prediction outcomes.

Shreyas Rajesh et al. [37] used Extra Trees Classifier, XGBoost, and Support Vector Machine (SVM) for churn modeling. The study found that these algorithms performed best with an 80:20 dataset distribution, achieving average AUC scores of 0.843 for Extra Trees, 0.787 for XGBoost, and 0.735 for SVM, while also minimizing false negatives. The paper highlights the need for a methodical churn prediction model to effectively monitor customers' churn behavior in the telecommunication industry, indicating a challenge in achieving desired performances in classifiers.

Sharmila Kishore et al.[38] used classification techniques such as Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree Classifier to analyze customer churn data in the telecom industry. The Random Forest classifier achieved an accuracy of 99%, with precision and recall both at 99%, resulting in an overall accuracy of 99.09% for the classification model. The paper does not explicitly mention any limitations or constraints associated with the customer churn prediction system developed using machine learning techniques in the telecom sector.

Daisy Reneilwe et al. [39] experimented with four algorithms—Logistic Regression, Random Forest, Support Vector Machines, and Extreme Gradient Boosting (XGBoost)—for churn prediction. The Random Forest algorithm yielded the best results, achieving an accuracy rate of 80%. The churn prediction model developed in this study also achieved an Area Under Curve (AUC) value of 84%, indicating strong performance.

Joydeb Kumar et al. [40] utilized various machine learning models and data transformation methods to enhance customer churn prediction in the telecommunications industry. Feature selection was performed using a univariate technique, and the best hyperparameters were selected through a grid search method to optimize the models. The study reports a 26.2% improvement in AUC and a 17% improvement in F-measure, attributing these gains to the application of data transformation methods and feature selection during the training of the churn prediction model.

Teuku Rizky Noviandy [41] evaluated five machine learning models—Naïve Bayes, Random Forest, AdaBoost, XGBoost, and LightGBM—for predicting customer churn in the telecommunications industry, following data preprocessing. Among these, LightGBM achieved the highest performance, with an accuracy of 80.70%, precision of 84.35%, recall of 90.54%, and an F1-score of 87.34, making it the most effective model for churn prediction.

Wee How Khoh [42] proposed a customer churn prediction system that employs three base learners—KNN, CatBoost, and Random Forest—within an ensemble model. The hyperparameters of these learners are tuned using grid search, and weighted soft voting is applied to combine the predictions based on the importance of each base learner. The system demonstrated strong performance with an accuracy of 84% and an F1 score of 83.42% on a real-world database, outperforming existing customer churn prediction systems, including both machine learning and deep learning models.

Gap Analysis and Recommendations

While reviewing the literature from the aforementioned angles, the following critical gaps have come to light, demanding attention while developing models for churn prediction.

- **Limited availability of up-to-date public datasets -** Most of the existing datasets are rather old or private. The former issue hampers the construction of churn prediction models with up-to-date features. The latter issue does not allow the

construction of replicable models, therefore affecting comparisons across studies.

- **Lack of consensus for feature-set -** Some studies suggest using behavioral features. In contrast, other studies suggest using customer feedback and network features while developing a churn prediction model.
- **Lack of consensus for classifiers -** There is a lack of consensus on the adoption of classifiers as there is a wide range that has been used in literature. While some studies suggest using simple classifiers such as regression and decision trees, other studies show ensembles and deep learning models have better performance. Also, the models often lack generalization across industry domains.
- **Drawbacks of traditional evaluation metrics –** Evaluation metrics such as accuracy, precision, recall and the ROC curve do not embed information about individual customer profitability. Acknowledging that not all customers hold equal value, these traditional machine learning-based metrics are insufficient to evaluate churn prediction models. This is because, for example, they do not consider the proportionate loss in profits due to distinct individual churn.
- **The tradeoff between model performance and its explainability -** Ensembles and deep learning-based approaches have proved helpful for building highly performant models. However, they lack interpretability and transparency. Therefore, further research is needed to develop explainable churn prediction methods, techniques, and models.

Here are refined recommendation for addressing the identified gaps:

- **Creation of novel public datasets -** To prioritise developing up-to-date, high-quality public datasets, companies should focus on data anonymization, ensuring privacy and compliance with relevant data protection laws and regulations. Establish clear data governance practices to ensure accountability, transparency, and responsible use of the shared datasets. Also, provide detailed documentation accompanying the dataset, including information on data sources, processing methods, and any transformations applied.
- **Integration of feature-sets -** While modelling churn prediction, it is recommended i) To combine behavioral features with demographics and ii) to incorporate information about social interactions and communication graphs of customers to model the influence of social circles iii) to consider customer feedback and perception on product or services to capture customers experiences, concerns, and their level of satisfaction. Analyzing feedback allows the identification of specific issues or areas of improvement.
- **Prioritise key notions on machine learning -** Instead of building a consensus on classifiers, researchers should focus on selecting the appropriate classifiers according to the underlying data, size, and shape. Similarly, they should prioritise the notions of model generalizability, the issue of the curse of dimensionality, underfitting/ overfitting and the issues behind class imbalance.

6. CONCLUSION

This research delves into the application of machine learning techniques for predicting customer churn in the OTT industry. By exploring various approaches such as supervised learning, unsupervised learning, and reinforcement-based systems, it highlights the potential of cognitive technologies in analyzing diverse data types to enhance churn prediction accuracy. This study provides a foundational framework for leveraging machine learning models in the OTT sector, offering valuable insights for addressing complex challenges related to customer retention. These findings underscore the importance of adopting advanced machine learning models to drive strategic decision-making and ensure sustainable growth in the competitive OTT landscape.

REFERENCES

[1] K. Khadka and S. Maharjan, ‘‘Customer satisfaction and customer loyalty,’’ Centria Univ. Appl. Sci. Pietarsaari, vol. 1, no. 10, pp. 58–64,2017.

[2] J. Ahn, J. Hwang, D. Kim, H. Choi, and S. Kang, ‘‘A survey on churn analysis in various Bus. Domains,’’ IEEE Access, vol. 8, pp. 220816–220839, 2020.

[3] J. N. Sheth and C. Uslay, ‘‘Creating enduring customer value,’’ J. Creating Value, vol. 8, no. 2, pp. 241–252, Nov. 2022.

[4] M.A.H. Farquad, Vadlamani Ravi, S. Bapi Raju “Churn prediction using comprehensible support vector machine: An analytical CRM application”, Applied Soft Computing 19 (2014) 31–40.

[5] 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,’’ IEEE Access, vol. 11, pp. 29759–29768, 2023.

[6] N. Edwine, W. Wang, W. Song, and D. Ssebuggwawo, ‘‘Detecting the risk of customer churn in telecom sector: A comparative study,’’ Math. Problems Eng., vol. 2022, pp. 1–16, Jul. 2022.

[7] L. C. Cheng, C.-C. Wu, and C.-Y. Chen, ‘‘Behavior analysis of customer churn for a customer relationship

system: An empirical case study,'' J. Global Inf. Manage., vol. 27, no. 1, pp. 111–127, Jan. 2019.

[8] A. Somosi, A. Stiassny, K. Kolos, and L. Warlop, ''Customer defection due to service elimination and post-elimination customer behavior: An empirical investigation in telecommunications,'' Int. J. Res. Marketing, vol. 38, no. 4, pp. 915–934, Dec. 2021.

[9] W. Soliman and T. Rinta-Kahila, ''Toward a refined conceptualization of is discontinuance: Reflection on the past and a way forward,'' Inf. Manage., vol. 57, no. 2, 2020, Art. no. 103167.

[10] H. Sebastian and R. Wagh, ''Churn analysis in telecommunication using logistic regression,'' Oriental J. Comput. Sci. Technol., vol. 10, no. 1, pp. 207–212, Mar. 2017.

[11] H. Lee, Y. Lee, H. Cho, K. Im, Y.S. Kim, "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model", Decision Support System 52 (2011) 207–216.

[12] H. Jain, A. Khunteta, and S. Srivastava, ''Telecom churn prediction and used techniques, datasets and performance measures: A review,'' Telecommun. Syst., vol. 76, no. 4, pp. 613–630, Apr. 2021.

[13] Kriti, "Customer churn: A study of factors affecting customer churn using machine learning" Iowa State University Capstones, Theses and Dissertations, March 2019.

[14] Essam Abou el Kassem, Shereen Ali Hussein, Alaa Mostafa Abdelrahman, Fahad Kamal Alsheref. "Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content" IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, November 5, 2020.

[15] I Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, Pratyush Sethi. "Customer churn prediction system: a machine learning approach" in Springer.

[16] Saran Kumar A., Chandrakala D. "A Survey on Customer Churn Prediction using Machine Learning Techniques" International Journal of Computer Applications (0975 – 8887) Volume 154 – No.10, November 2016.

[17] Pradeep B, Sushmitha Vishwanath Rao, Swati M Puranik, Akshay Hegde "Analysis of Customer Churn prediction in Logistic Industry using Machine Learning" International Journal of Scientific and Research Publications, Volume 7, Issue 11, November 2017 ISSN 2250-3153.

[18] Koen W. De Bock, Dirk Van den Poel, "Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models", Expert Systems with Applications 39 (2012) 6816–6826.

[19] S. H. Iranmanesh, M. Hamid, M. Bastan, G. Hamed Shakouri, and M. M. Nasiri, ''Customer churn prediction using artificial neural network: An analytical CRM application,'' in Proc. Int. Conf. Ind. Eng. Oper. Manag., 2019, pp. 23–26.

[20] J. Bojei, C. C. Julian, C. A. B. C.Wel, and Z. U. Ahmed, ''The empirical link between relationship marketing tools and consumer retention in retail marketing,'' J. Consum. Behaviour, vol. 12, no. 3, pp. 171–181, May 2013.

[21] K. Ljubičić, A. Merćep, and Z. Kostanjčar, ''Churn prediction methods based on mutual customer interdependence,'' J. Comput. Sci., vol. 67, Mar. 2023, Art. no. 101940.

[22] D. L. García, À. Nebot, and A. Vellido, ''Intelligent data analysis approaches to churn as a business problem: A survey,'' Knowl. Inf. Syst., vol. 51, no. 3, pp. 719–774, Jun. 2017.

[23] J. Kandampully, T. Zhang, and A. Bilgihan, ''Customer loyalty: A review and future directions with a special focus on the hospitality industry,'' Int. J. Contemp. Hospitality Manage., vol. 27, no. 3, pp. 379–414, Apr. 2015.

[24] H.Wetzel, C. Haenel, and A. C. Hess, ''Handle with care! Service contract termination as a service delivery task,'' Eur. J. Marketing, vol. 56, no. 12, pp. 3169–3196, Nov. 2022.

[25] D. L. García, A. V. Alcacena, and M. À. Castells, ''Customer continuity management as a foundation for churn data mining,'' 2007.

[26] D. Ismanova, ''Students' loyalty in higher education: The mediating effect of satisfaction, trust, commitment on student loyalty to alma mater,'' Manage. Sci. Lett., vol. 9, no. 8, pp. 1161–1168, 2019.

[27] H. Ribeiro, B. Barbosa, A. C. Moreira, and R. G. Rodrigues, ''Determinants of churn in telecommunication services: A systematic literature review,'' Manage. Rev. Quart., vol. 1, pp. 1–38, Feb. 2023.

[28] A. Ben, ''Enhanced churn prediction in the telecommunication industry,'' SSRN Electron. J., vol. 8, no. 2, pp. 6–15, 2020.

[29] P. S. H. Tan, Y. O. Choong, and I.-C. Chen, ''The effect of service quality on behavioral intention: The

mediating role of student satisfaction and switching barriers in private universities,'' J. Appl. Res. Higher Educ., vol. 14, no. 4, pp. 1394–1413, Dec. 2022.

[30] E. Ghazali, B. Nguyen, D. S. Mutum, and A. A. Mohd-Any, ''Constructing online switching barriers: Examining the effects of switching costs and alternative attractiveness on e-store loyalty in online pure-play retailers,'' Electron. Markets, vol. 26, no. 2, pp. 157–171, May 2016.

[31] Amiya, Ranjan, Panda., Manoj, Kumar, Mishra., Dvij, Kalsi., Kaushik, Jyoti, Bhuyan., Soumyadeep, Saha., Kaustubh, Jyoti, Bhuyan. (2024). 1. Classification of Customer Churning based on OTT platform data. doi: 10.1109/esic60604.2024.10481640.

[32] Tasneem, Qaraeen., Nora, Qaqour., Sameh, Taqatqa. (2024). 8. Predictive Customer Analytics: Machine Learning for Churn Prediction and Retention. doi: 10.59994/ajbtme.2024.1.11.

[33] Victor, Chang., Karl, Hall., Qianwen, Xu., Folakemi, Ololade, Amao., Meghana, Ashok, Ganatra., Vladlena, Benson. (2024). 6. Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models. Algorithms, doi: 10.3390/a17060231.

[34] Muteb, Alotaibi., Mohd, Anul, Haq. (2024). 10. Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods. Engineering, Technology & Applied Science Research, doi: 10.48084/etasr.7480

[35] Opara, John, Ogbonna., Gilbert, I.O., Aimufua., Muhammad, Umar, Abdullahi., Suleiman, Abubakar. (2024). 17. Churn Prediction in Telecommunication Industry: A Comparative Analysis of Boosting Algorithms. doi: 10.4314/dujopas.v10i1b.33.

[36] Xiayu, Li., Tianyi, Yang., Xu, Zhan., Yanxin, Shi., Huixiang, Li. (2024). 18. Utilizing Data Science and AI for Customer Churn Prediction in Marketing. Journal of Theory and Practice of Engineering Science, doi: 10.53469/jtpes.2024.04(05).10.

[37] Muhammad, Maulana, Sidiq., Dyah, Anggraini. (2023). 9. Analysis and Classification of Customer Churn Using Machine Learning Models. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), doi: 10.29207/resti.v7i6.4933.

[38] Manal, Loukili., F., Messaoudi., Raouya, El, Youbi. (2023). 7. Implementation of Machine Learning Algorithms for Customer Churn Prediction. doi: 10.61186/jist.34208.11.43.196.

[39] Daisy, Reneilwe, Chabumba., Ashwini, Jadhav., Ritesh, Ajoodha. (2021). 34. Predicting Telecommunication Customer Churn using Machine Learning Techniques. doi: 10.1201/9781003202240-98

[40] Joydeb, Kumar, Sana., Mohammad, Zoynul, Abedin., M., Sohel, Rahman., Md., Saidur, Rahman. (2022). 37. A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection. PLOS ONE, doi: 10.1371/journal.pone.0278095

[41] Teuku, Rizky, Noviandy., Ghalieb, Mutig, Idroes., Irsan, Hardi., Mohd, Afjal., Samrat, Ray. (2024). 44. A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry. Infolitika Journal of Data Science, doi: 10.60084/ijds.v2i1.199.

[42] Wee, How, Khoh., Ying, Han, Pang., Shih, Yin, Ooi. (2023). 48. Predictive Churn Modeling for Sustainable Business in the Telecommunication Industry: Optimized Weighted Ensemble Machine Learning. Sustainability, doi: 10.3390/su15118631

[43] Liwei, Chen. (2023). 16. Machine Learning-based Analysis and Prediction of Telecoms Customer Churn. doi: 10.1109/icaml60083.2023.00032.

[44] Navienkumar, R., Lalithamani, N. (2023). 25. Machine Learning Methods for Predictive Customer Churn Analysis in the Telecom Industry. doi: 10.1109/icccnt56998.2023.10306395.

[45] Angela, Yi, Wen, Chong., Khai, Wah, Khaw., Wai, Chung, Yeong., Wen, Xu, Chuah. (2023). 47. Customer Churn Prediction of Telecom Company Using Machine Learning Algorithms. JOURNAL OF SOFT COMPUTING AND DATA MINING, doi: 10.30880/jscdm.2023.04.02.001.

[46] Abdelrahim, Kasem, Ahmad., Assef, Jafar., Kadan, Aljoumaa. (2019) Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, doi: 10.1186/S40537-019-0191-6.

[47] Alae, Chouiekh., El, Hassane, Ibn, El, Haj. (2020). 66. Deep Convolutional Neural Networks for Customer Churn Prediction Analysis. International Journal of Cognitive Informatics and Natural Intelligence, doi: 10.4018/IJCINI.2020010101.

[48] Shreyas, Rajesh, Labhsetwar. (2020). 74. Predictive analysis of customer churn in telecom industry using supervised learning. doi: 10.21917/IJSC.2020.0291.

[49] Sharmila, Kishor, Wagh., Aishwarya, A., Andhale., Kishor, S., Wagh., J., Pansare., Sarita, P., Ambadekar., S.,

H., Gawande. (2023). 76. Customer Churn Prediction in Telecom Sector using Machine Learning Techniques. Results in control and optimization, doi: 10.1016/j.rico.2023.100342.

[50] S. Mitrović and J. De Weerdt, ''Churn modeling with probabilistic meta paths-based representation learning,'' Inf. Process. Manage., vol. 57, no. 2, Mar. 2020, Art. no. 102052.

[51] A. Dursun and M. Caber, ''Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis,'' Tourism Manage. Perspect., vol. 18, pp. 153–160, Apr. 2016.

[52] K. Georgiou and A. Chasapis, ''Novel time slicing approach for customer defection models in e-commerce: A case study,'' Data Sci. Manage., vol. 5, no. 3, pp. 149–162, Sep. 2022.

[53] G. Mena, K. Coussement, K. W. De Bock, A. De Caigny, and S. Lessmann, ''Exploiting time-varying RFM measures for customer churn prediction with deep neural networks,'' Ann. Operations Res., vol. 1, pp. 1–23, Mar. 2023.

[54] T. Gattermann-Itschert and U. W. Thonemann, ''Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests,'' Ind. Marketing Manage., vol. 107, pp. 134–147, Nov. 2022.

[55] M. Mirkovic, T. Lolic, D. Stefanovic, A. Anderla, and D. Gracanin, ''Customer churn prediction in B2B non-contractual Bus. Settings using invoice data,'' Appl. Sci., vol. 12, no. 10, p. 5001, May 2022.

[56] B. Janssens, M. Bogaert, A. Bagué, and D. Van den Poel, ''B2Boost: Instance-dependent profit-driven modelling of B2B churn,'' Ann. Operations Res., vol. 1, pp. 1–27, Mar. 2022.

[57] S. Maldonado, G. Domínguez, D. Olaya, andW. Verbeke, ''Profit-driven churn prediction for the mutual fund industry: A multi segment approach,''Omega, vol. 100, Apr. 2021, Art. no. 102380.

[58] S. Khodabandehlou and M. Zivari Rahman, ''Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior,'' J. Syst. Inf. Technol., vol. 19, no. 1, pp. 65–93, Mar. 2017.

[59] M. Y. Smaili and H. Hachimi, ''New RFM-D classification model for improving customer analysis and response prediction,'' Ain Shams Eng. J., vol. 14, no. 12, Dec. 2023, Art. no. 102254.

[60] K. Chen, Y.-H. Hu, and Y.-C. Hsieh, ''Predicting customer churn from valuable B2B customers in the logistics industry: A case study,'' Inf. Syst. E-Business Manage., vol. 13, no. 3, pp. 475–494, Aug. 2015.

[61] S. Peker, A. Kocyigit, and P. E. Eren, ''LRFMP model for customer segmentation in the grocery retail industry: A case study,'' Marketing Intell. Planning, vol. 35, no. 4, pp. 544–559, May 2017.

[62] T. Gattermann-Itschert and U.W. Thonemann, ''How training on multiple time slices improves performance in churn prediction,'' Eur. J. Oper. Res., vol. 295, no. 2, pp. 664–674, Dec. 2021.

[63] A. Perišic and M. Pahor, ''RFM-LIR feature framework for churn prediction in the mobile games market,'' IEEE Trans. Games, vol. 14, no. 2, pp. 126–137, Jun. 2022.

[64] S. Mitrović, B. Baesens, W. Lemahieu, and J. De Weerdt, ''On the operational efficiency of different feature types for Telco churn prediction,'' Eur. J. Oper. Res., vol. 267, no. 3, pp. 1141–1155, Jun. 2018.

[65] O. M. Mirza, G. Jose Moses, R. Rajender, E. Laxmi Lydia, S. Kadry, C. Me-Ead, and O. Thinnukool, ''Optimal deep canonically correlated autoencoder-enabled prediction model for customer churn prediction,'' Comput., Mater. Continua, vol. 73, no. 2, pp. 3757–3769, 2022.