2024

A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners

Awais Manzoor
Technological University Dublin

M. Atif Qureshi
Technological University Dublin, atif.qureshi@tudublin.ie

Etain Kidney
Technological University Dublin

See next page for additional authors

Funder: Irish Research Council

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 March 2024, accepted 13 May 2024, date of publication 16 May 2024, date of current version 24 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3402092

SURVEY

A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners

AWAIS MANZOOR[1,2,3], M. ATIF QURESHI[1,2], ETAIN KIDNEY[2], AND LUCA LONGO[3]

[1]ADAPT Centre, eXplainable Analytics Group, Technological University Dublin, Dublin 2, D02 HW71 Ireland
[2]Faculty of Business, Technological University Dublin, Dublin 2, D02 HW71 Ireland
[3]Artificial Intelligence and Cognitive Load Research Laboratories, Technological University Dublin, Dublin 7, D07 H6K8 Ireland

Corresponding author: Awais Manzoor (malik.awaismanzoor@gmail.com)

This work was supported by the Irish Research Council under Award GOIPG/2021/1354; and in part by the Science Foundation Ireland through the ADAPT SFI Research Centre, Technological University Dublin, under Grant 13/RC/2106_P2.

ABSTRACT Due to market deregulation and globalisation, competitive environments in various sectors continuously evolve, leading to increased customer churn. Effectively anticipating and mitigating customer churn is vital for businesses to retain their customer base and sustain business growth. This research scrutinizes 212 published articles from 2015 to 2023, delving into customer churn prediction using machine learning methods. Distinctive in its scope, this work covers key stages of churn prediction models comprehensively, contrary to published reviews, which focus on some aspects of churn prediction, such as model development, feature engineering and model evaluation using traditional machine learning-based evaluation metrics. The review emphasises the incorporation of features such as demographic, usage-related, and behavioural characteristics and features capturing customer social interaction and communications graphs and customer feedback while focusing on popular sectors such as telecommunication, finance, and online gaming when producing newer datasets or developing a predictive model. Findings suggest that research on the profitability aspect of churn prediction models is under-researched and advocates using profit-based evaluation metrics to support decision-making, improve customer retention, and increase profitability. Finally, this research concludes with recommendations that advocate the use of ensembles and deep learning techniques, and as well as the adoption of explainable methods to drive further advancements.

INDEX TERMS Churn prediction, machine learning, artificial intelligence, business decision making, customer defection, marketing analytics, business intelligence.

I. INTRODUCTION

The challenge to remain competitive and maintain good customer relationships often leads businesses to adopt new methods and technologies to increase revenue. However, acquiring new customers is generally more expensive than retaining existing ones [1]. Consequently, this has led to the development of customer churn prediction and retention strategies [2], especially focused on retaining customers who offer a higher rate of return on investment [3]. Retaining customers can also reduce the cost of marketing for their acquisition, as they may spread positive word of mouth about a business [4]. Customer churn is a significant challenge businesses face when customers abandon or switch to other service providers. Different industries have varying criteria to classify churners as a rule of thumb. For instance, experts in the banking sector consider a customer a churner if their annual transactions and average annual account balance decrease by 30% [5]. Similarly, Edwine et al. [6] define a customer as a churner if no activity for 90 consecutive days

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

is demonstrated. Churn can occur due to reasons, such as dissatisfaction with service quality [7] or as a response to price increases [8] in the telecommunication sector. To minimise churn rates, businesses need to anticipate customer churn by analysing customer behaviour and addressing potential underlying causes of dissatisfaction. Research has shown that even a modest increase in customer retention rate can lead to substantial increases in net present value for companies, with software and advertising companies experiencing up to 35% and 95% increases, respectively [2].

Customer churn can be divided into two categories: *intentional* and *involuntary*. Intentional churn occurs when customers voluntarily discontinue using a service because of dissatisfaction or a change in circumstances [9]. They may switch to another provider that offers alternate plans, better service quality, or both. To retain customers, businesses can address the issue by offering attractive plans and ensuring high-quality service. On the other hand, involuntary churn happens when the business stops providing service due to unpaid bills or breaches of terms and conditions [10]. There are two sub-categories within intentional churn: *deliberate* and *incidental* churn. Deliberate churn occurs when customers switch to another provider due to dissatisfaction with the product or service to obtain a better alternative. This dissatisfaction may stem from technical issues such as poor service quality, outdated services, bad customer service experience, poor coverage, or financial concerns such as expensive plans. On the other hand, incidental churn happens when customers stop using the service due to changes in their circumstances, such as relocating to another city where the service is unavailable, switching to another job that limits their use of a specific service provider, or the services becoming unaffordable for the customer. The primary goal of churn prediction is to anticipate deliberate churn since involuntary customers who breach terms and conditions or fail to pay bills are already known to the business [11]. Incidental churn accounts for a small portion of churn, and it is difficult to predict since even the customers themselves may not foresee changes such as a change of place or job before a specific change in reality. Additionally, knowing incidental churn in advance is of little value since retaining those customers becomes unavoidable due to reasons beyond the products and services offered by the business [12]. In the next section, an in-depth discussion of the need for churn prediction in the business context is presented. Technical readers interested in developing churn prediction models can skip it.

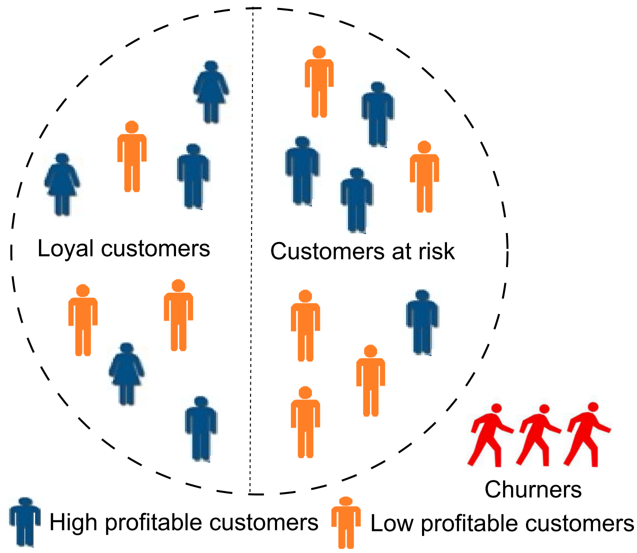


FIGURE 1. An illustration of loyal customers and churners.

A. NEED OF CHURN PREDICTION IN BUSINESS

In a highly competitive global market, the economic growth of businesses relies on both prolonging the average lifespan of existing customers and increasing their consumption. To achieve customer retention, it is crucial to understand the factors contributing to customer loyalty development and to predict customers' churning intentions, thus developing effective retention campaigns. By comprehending this mechanism, businesses can develop strategies to retain existing customers and build strong loyalty bonds with them, which is less costly and more effective than attracting new customers [13]. Losing profitable customers can be expensive, which has led to a shift in the market paradigm from an emphasis on acquiring new customers to retaining existing ones. Additionally, churn influence propagates in the social circle as customer mutual interdependence exists, and when churn happens, it also increases churn probability in the social circle [14]. Despite businesses' efforts to build customer loyalty, they face challenges due to customers' exposure to competing advertised offers and market awareness. To ensure the sustainability and growth of the business, it is crucial to understand the mechanism of developing customer loyalty bonds [15]. Customers are the primary assets of any business, and companies are responsible for defining and implementing policies that prolong their lifetime. Prolonging customers' lifetimes should be the primary focus of business firms, as it determines their future commercial direction [16]. Forcing customers to upgrade their purchases without fulfilling their real needs can lead to high customer abandonment and lower customer satisfaction, ultimately damaging a company's reputation [2].

However, terminating unprofitable customers may have an unintended negative effect on the loyalty of retained valuable customers [17]. Building customer loyalty bonds requires a systematic approach to evaluating their evolution over time and anticipating possible signs of defection. Customer continuity management is an organisational approach to retaining and acquiring customers [18]. A customer continuity management model should provide a comprehensive critical review of all aspects that might affect the development of customer loyalty bonds. While customer loyalty bonds and lifetime are directly related to customer satisfaction

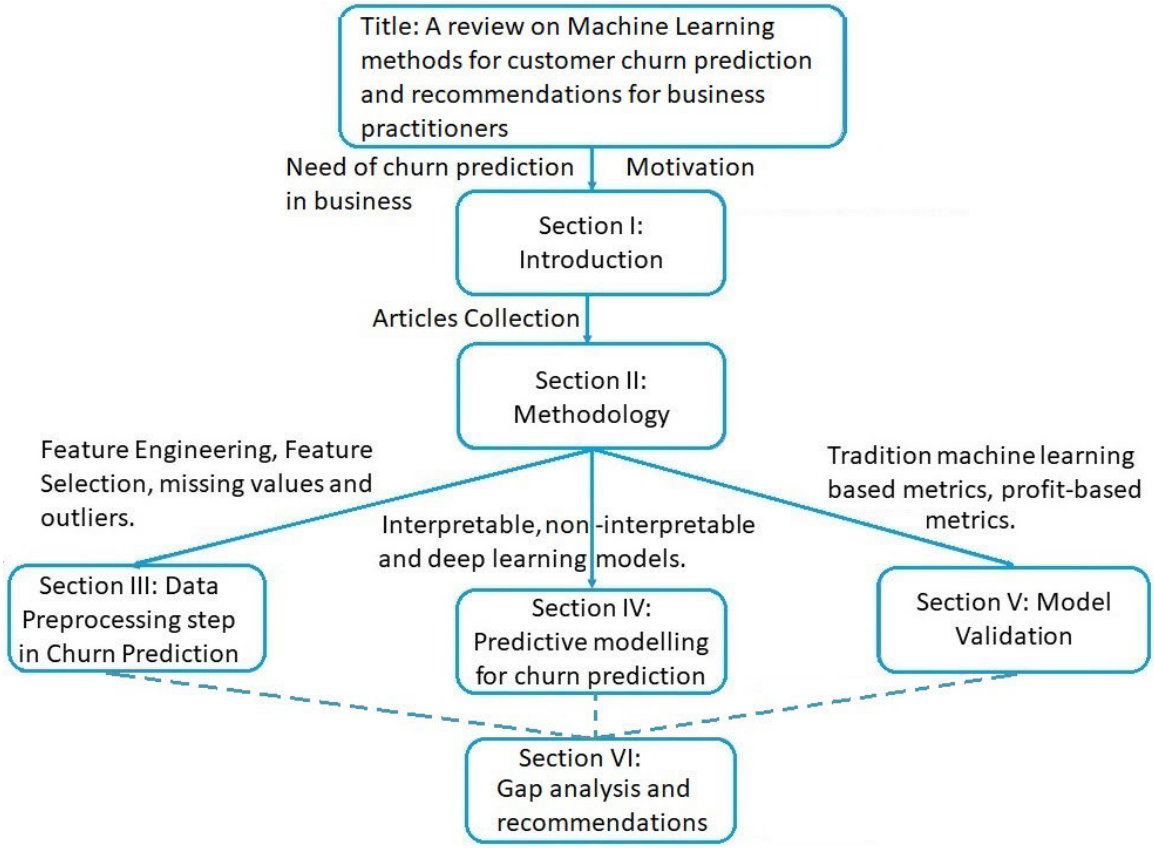


FIGURE 2. **Literature review document outline.**

in terms of service quality [19], it is important to note that service quality does not always guarantee customer loyalty. For instance, customers may switch to another telecommunication operator despite receiving good service or prefer to fly with specific airlines despite experiencing flight delays.

Limiting customers' freedom to switch through contractual restrictions can be considered an alternative approach to managing the loyalty-satisfaction relationship [20], [21], [22]. However, such restrictions may negatively affect customer loyalty [23] and may not be viable in the long term. Therefore, switching barriers should be built based on customer perception [24]. Similarly, due to the heterogeneous nature of customer perception and service requirements, it is impossible to establish a loyalty bond with all customers. Customers usually do not churn due to a single dissatisfaction scenario; there are usually multiple instances of dissatisfaction before a customer ceases transactions with an organisation [25]. Consequently, dissatisfied customers will always exist, so businesses focus on developing retention campaigns by improving their offers and implementing effective switching barriers. Figure 1 illustrates the customer churn lifecycle, categorising customers into loyal, at-risk, and churned segments. It underscores the importance of stratifying customers based on profitability, discerning between highly and non-highly profitable segments. The fundamental objective of customer churn prediction is to mitigate churn, especially among highly profitable customers, directing retention efforts toward those substantially impacting profitability.

Throughout a lifespan, external factors such as competitor offers, new product launches, and customer awareness may influence customers' expectations and satisfaction [26]. As a result, businesses must adapt their policies, and since obtaining opinions from all customers can be costly, they often tend to analyse representative samples. When developing retention strategies, businesses should consider the expected lifetime value of each customer and compare it to the cost of the retention efforts. Retaining unprofitable customers who do not justify the cost of the retention campaign may not be cost-effective for the business. Therefore, it is essential to balance the costs and benefits of retention campaigns to make informed decisions. It is more beneficial for businesses to focus their resources on retaining a targeted group of customers rather than attempting to increase the overall retention rate for all existing customers [27].

B. MOTIVATION

While substantial interest in churn prediction has led to numerous review articles on the topic, a notable gap remains in the availability of comprehensive guidelines for businesses aiming to develop effective churn prediction models. This knowledge deficit underscores the need for a thorough survey providing insights into the end-to-end customer churn prediction pipeline. This imperative set the stage for our comprehensive survey, with a primary research question guiding our exploration:

> "How can businesses, particularly telecommunications, finance, and online gaming sectors, leverage insights and methods from current Machine Learning research to predict customer churn?"

To elucidate the motivation behind our survey and the research question, Table 2 briefly summarises the focus of review articles published between January 2015 and January 2024, offering a comparative analysis with our survey. The table distinctly identifies a crucial gap in the literature related to feature selection strategies, the incorporation of profit-based metrics for model validation, sampling strategies, and contextualisation of the trend of predictive modelling. This survey aims to fill these gaps and provide a holistic understanding of the customer churn prediction landscape.

As shown in Figure 2, the review is structured as follows. The methodology adopted for this literature review is detailed in Section II. Section III focuses on describing the steps often taken to pre-process data for churn prediction, followed by a summary. Section IV is dedicated to describing the ways predictive models can be trained for churn prediction. Section V focuses on the validation strategies for such prediction models. Finally, Section VI synthesises the gaps and offers recommendations contributing to the body of knowledge.

II. METHODOLOGY

A three-step systematic literature review was employed to provide an impartial and objective assessment of the present state-of-the-art in churn prediction and future applications of machine learning within the context of churn prediction. Firstly, the scope of the study was narrowed down to include articles using machine learning to predict customer churn. Secondly, relevant articles were identified and selected from chosen databases by selecting key search terms and the query refinement. Lastly, the results and key findings of relevant studies were analysed and synthesised.

A. ARTICLES COLLECTION

To ensure a comprehensive and diverse topic coverage, we conducted an initial literature scan using the primary terms "Customer Prediction" and "Machine Learning" on Web of Science (WoS), DBLP, Scopus and Google Scholar to collect initial keywords. Following this, we expanded the initial search terms by including additional keywords: "Customer defection", "Customer turnover", "Customer switching", and "Customer abandonment". These keywords were combined with "Machine Learning" or "Artificial Intelligence" in the final search query.

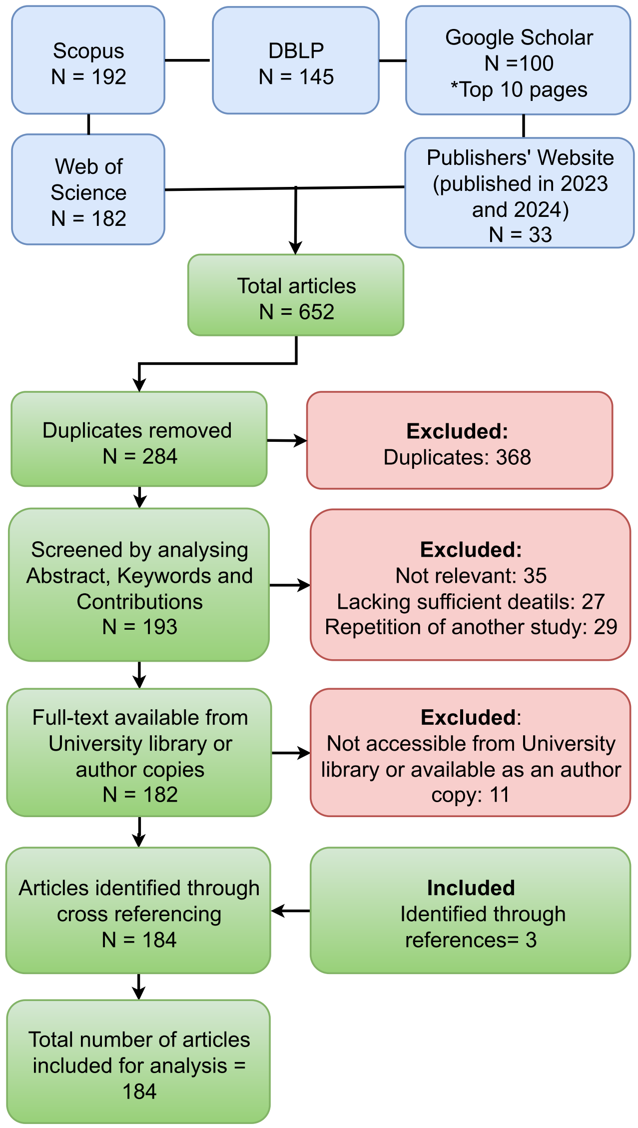


FIGURE 3. Flowchart illustrating article count and inclusion/exclusion criteria in the study.

Finally, the search was conducted on January 25, 2024, and limited to articles published between 2015[1] and January-2024, using the following Boolean search query: ("Customer churn" OR "Customer turnover" OR "Customer abandonment" OR "Customer defection" OR "Customer attrition" OR "Customer switching") AND ("Machine Learning" OR "Artificial Intelligence")

At first, the search strategy retrieved 182 articles from WoS and 192 from Scopus. These articles were then filtered to include only those published in journals. Following that, a similar search strategy on Google Scholar and DBLP was performed, considering articles from the first ten pages of

[1]The choice of the timeframe, from 2015 onwards, is based on the observation that research on customer churn prediction using machine learning gained significant attention during this decade (refer to Figure 4b).

TABLE 1. Inclusion and exclusion criteria; here, WoS and GS refer to WoS and google scholar, respectively.

Database	Inclusion Criteria	Exclusion Criteria
WoS, Scopus, publisher's website, GS, DBLP	Articles relevant to churn prediction using AI or ML techniques	Articles not relevant to customer churn prediction or do not utilise AI or ML techniques
WoS, Scopus, publisher's website (2023 and 2024 only), GS, DBLP	Articles published between 2015 and Jan-2024	Articles published before 2015.
WoS, Scopus, publisher's website, GS DBLP	Relevant articles providing sufficient details and not duplicating other studies	Articles lacking adequate details, impeding scientific reproducibility, or duplicating other studies
(Filtered Literature)	Relevant pre-2015 articles discovered through 2015 to Jan-2024 references	Articles published before 2015 but discovered as not highly relevant

the search results of Google Scholar.[2] This search yielded 100 articles from Google Scholar and 145 from DBLP. We also searched the websites of different publishers, including MDPI, IEEE Explore, and Springer, for articles published in 2023 and 2024 and found 33 additional articles.[3] Duplicate articles were removed, leading to an overall of 181 articles.

Next, a thorough analysis of the articles' abstracts, keywords, and key contributions was performed. Only those articles that met the inclusion criteria as shown in Table 1, and were accessible as full-rom Technological University Dublin, Ireland[4] were included in this study.

During the thorough analysis, three additional articles published before 2015 were discovered and were deemed relevant. These articles were included in this study because they provided unique and valuable insights into profitability-based metrics in the context of customer churn prediction. Furthermore, the articles that replicated existing studies or those that were not centred on machine learning were excluded. Also, studies lacking adequate details or scientific reproducibility,[5] were excluded from the selection process.

After careful analysis, a set of 185 articles that met the inclusion criteria was selected. Figure 3 provides a flowchart detailing the number of articles and their inclusion and exclusion criteria. Additionally, 27 supportive articles were included to offer a comprehensive narrative and business context. Ultimately, this literature review condensed to a count of 212 articles, and Figures 4a and 4b show distributional statistics concerning the total number of articles.

[2]While [28] suggests the first four pages of the search retrieval are relevant, we expanded our search to the first ten pages for thoroughness and completeness).

[3]The motivation behind exploring articles published in 2023 on publishers' websites was to identify recently published articles that may not be indexed in databases.

[4]For inaccessible articles, we sought alternative sources like author copies and requested access from colleagues at other universities. However, despite these efforts, a few articles remained inaccessible and were consequently excluded.

[5]such as missing information on the number and types of features concerning private datasets or clear details on model implementation.

III. DATA PREPROCESSING STEPS IN CHURN PREDICTION

The pre-processing phase is the first step of the widely adopted end-to-end churn prediction modelling pipeline, as shown in Figure 5.

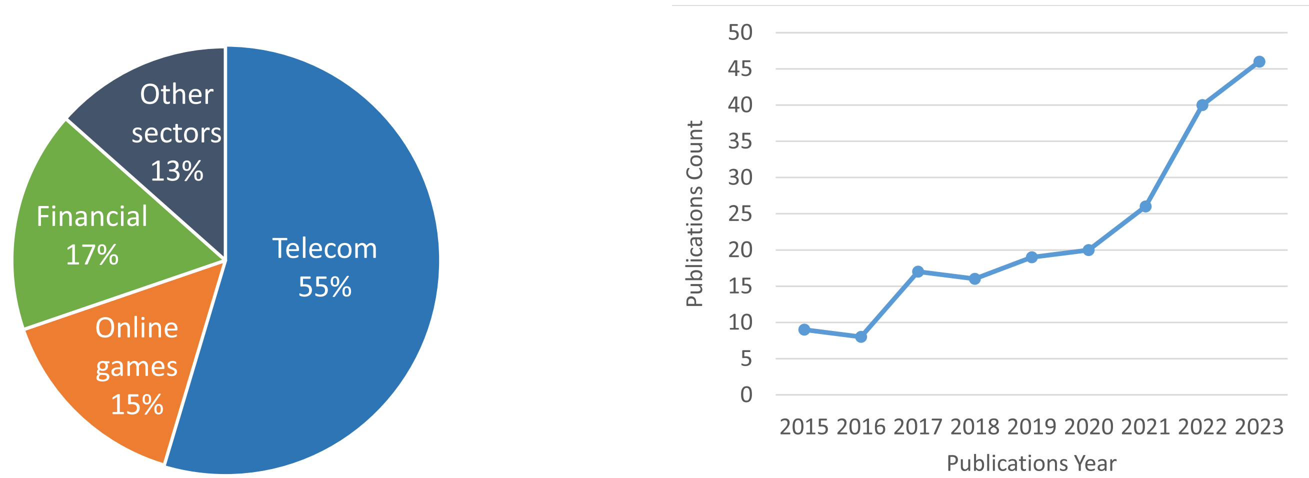
The customer data is typically stored in databases for churn prediction, a common practice in businesses. However, this data may contain noise or missing values, which can negatively affect the accuracy of the churn predictive model. Therefore, it is necessary to improve data quality through techniques like data cleaning, filtration, transformation, and reduction, all of which are part of the so-called pre-processing phase. Even though the attributes or features of a dataset may initially appear self-explanatory, and they can belong to different data types, such as textual or numbers. However, there is no standard way or one-size-fits-all approach to store data that meets all business needs [15].

Data storage primarily requires a database manager to interact with other departments and collect various forms of data, including textual, numerical, categorical, and geographical data [32]. It is the responsibility of data scientists to prepare the data required to build a predictive model, which includes data cleaning, handling missing values, and mapping data to high-level categories, such as geo-coordinates, among other tasks [33].

A. HANDLING MISSING VALUES, OUTLIERS AND CLASS IMBALANCE

Missing values and outliers can significantly affect churn prediction models. To handle missing values, researchers typically use imputation or removal techniques. Imputation methods such as mean, median, or mode [34] are typically applied to attributes with more than 5% missing values, while instances with less than 5% missing values can be dropped [35]. For instance, Azeem and Usman [36] removed features with the highest number of missing values from a telecommunications churn dataset, while Höppner et al. [37] removed observations with missing values.

Outliers or extreme values can also affect model performance and must be identified and handled. One common approach is removing outliers over three standard deviations from the mean [38]. Other techniques to detect and handle



(a) Sector-wise distribution of articles

(b) Year-wise Publication)

FIGURE 4. Yearly and sector-wise publications trend (Published until Dec-2023).

TABLE 2. Summary of important surveys on customer churn prediction and artificial intelligence.

Ref	Publication year	Sector(s)	Predictive models	Features Engineering	Feature selection	Sampling Strategies	Traditional evaluation metrics	Profitability based metrics	Remarks
[29]	2017	Telecom	Medium	Low	-	-	Low	-	Performed comparative analysis of techniques through implementation
[15]	2017	Multiple	High	Medium	Medium	-	Medium	-	Article published in 2017, and lacks sufficient discussion on deep learning, evaluation metrics and feature engineering
[2]	2020	Multiple	Low	High	High	Low	Low	-	Generally discussed multiple sectors, not focused any sector
[11]	2020	Telecom	Medium	Low	-	-	Medium	-	Comparative analysis of techniques through model implementation
[30]	2022	Multiple	High	-	-	High	Medium	-	Comparative analysis of predictive models with and without data sampling
[20]	2023	Telecom	-	High	-	-	-	-	The article is discussed from the perspective of churn indicators (feature leading to customer churning)
[31]	2023	Multiple	High	-	-	-	High	Medium	The survey is based on comparative analysis of classifiers mainly focusing on ensembles methods
Our	——	**Multiple**	High	High	High	High	High	High	Comprehensive survey of the literature covering key stages of churn prediction

outliers include data binning, imputation, and trimming [39]. In their study on customer churn prediction, Azeem and Usman [36] removed outliers with ±5 unit values from the entire dataset based on the business context. Similarly, researchers often filter call detail record (CDR) by dropping calls lasting less than 4 seconds, often considered unintentional [40].

Class imbalance is a significant factor that can impact the reliability of a classifier. In service-based industries, churn is rare, resulting in imbalanced datasets, and misclassification is more costly in such cases [41]. Unbalanced class distribution often leads to the underrepresentation of the minority class, resulting in the classifier being under-trained in the minority classes [6]. Resampling techniques like random oversampling [42] and undersampling strategies [35] are commonly used to address the class imbalance issue in churn prediction.

Random oversampling involves replicating instances of the minority class until the minority and majority classes are approximately balanced, while undersampling randomly removes instances from the majority class. In the studies [6], [43], [44], different undersampling strategies were applied to a telecommunication churn dataset. SMOTE (Synthetic Minority Oversampling Technique) oversampling has been employed in churn prediction problems in studies such as [45], [46], [47], [48], [49], and [50]. Reference [51] has employed SMOTE-nominal continuous (SMOTE-nc) to oversample churners. SMOTE generates synthetic instances of the minority class to alleviate the class imbalance issue.

Reference [52] proposed a bagging-based selective ensemble model to improve the predictive performance of customer churn on the imbalanced dataset. Their strategy exploits overproduction and chooses a strategy using a cost-weighted negative binomial distribution to generate random class ratios in the training subsets. Then, it uses cost-sensitive logistic regression with a lasso penalty to aggregate base classifiers. They compared their results with twelve state-of-the-art classification techniques on telecommunication datasets and showed that their method outperforms its counterparts on traditional and profit-based metrics. Similarly, [47] developed an ensemble by combining several classifiers (Neural Network-XGBoost-Adaboost, KNN-XGBoost-Adaboost, SVM-XGBoost-Adaboost, Random Forest-XGBoost-Adaboost, Logistic Regression-XGBoost-Adaboost, Decision Tree-XGBoost-Adaboost) and compared their performances on customer churn datasets with standalone base classifiers, showing that the ensemble of logistic regression, neural network, and AdaBoost has the best predictive performance. Reference [53] addressed the class imbalance in telecom churn datasets by the undersampling method through particle swarm optimisation, which provides an unbiased distribution of the training set to the ensemble approach based on genetic programming and AdaBoost.

Reference [54] applied different variations of generative adversarial networks (GAN) for oversampling of the minority class and compared the efficacy of the proposed oversampling techniques with several classifiers, demonstrating that oversampling through GAN improves predictive performance. Reference [55] applied both oversampling and undersampling on the financial dataset. First, the adaptive synthetic sampling approach (ADASYN) was applied for oversampling, and then the NearMiss[6] method was applied for undersampling. ADASYN added 5% artificial churn customers, and NearMiss removed 65% of the non-churn customers to balance the dataset. Pustokhina et al. [56] applied a multi-objective rain optimisation algorithm to determine the optimal sampling size for the synthetic minority oversampling technique (SMOTE). Similarly, [41] compared the performances of six oversampling techniques (SMOTE, mega-trend-diffusion function (MTDF), minimal redundancy maximal relevance (mRMR), ADASYN, majority-weighted minority oversampling technique (MWMOTE), immune centroids oversampling technique (ICOTE), and couples top-N reverse k-nearest neighbour (TRkNN)) on six imbalanced datasets and found that MTDF yielded the best performance to deal with the imbalance problem in the context of customer churn in the telecommunication sector.

B. FEATURE ENGINEERING

Feature engineering is the process of extracting relevant variables from raw data using domain knowledge, often depending on the business domain and the task context. In industries like telecommunications and gaming, churn prediction models often rely on customer usage behaviour, and the recent activity logs are commonly used indicators of churn behaviour [36], [58]. However, relying solely on short-term activity from logs may lead to misclassification [2] or determine churn behaviour too late to effectively persuade a customer to remain as such [59]. To address this, researchers have suggested the identification of other relevant features that are as important as customers' recent activity logs before churning or the utilisation of time series log data to identify potential churners [60], [61].

In financial industry, Ismail et al. associated churn with demographic features and company-customer relationship data, including gender, industry code, tenure, service suspension/resumption frequency, and average invoice [62]. M. Alizadeh et al. identified financial, behavioural and demographic features as predictors of customer churn in the banking sector [5]. Specifically, average account balance, customer relationship duration, and transaction frequency were identified as the most relevant financial variables. Likewise, behavioural features have also been recognised as important factors for churn predictions in various industries, including online gaming [59], [63], [64], telecommunication [65], [66] and the financial sector [67].

In online gaming, user behaviour based on in-game activities and game logs was exploited to predict customer churn [68], [69]. [68] claims that time investments in online gaming are also key indicators for predicting customer churn. L. Kim et al. observed that social relationships among game players and irregularities in time-spending have a direct relationship with customer churn, and particularly, the latter increases the probability of churn [70]. Kim et al. [63] observed that active duration, play count, win ratio and purchase count are also highly relevant features for churn prediction. Additionally, RFM-based behavioural features have been recognised as important churn predictors [71], [72] and an optimal method for analysing customer behaviour [45].

RFM-based features, which include recency, frequency, and monetary variables derived from invoice data, have been widely used. They offer valuable insights into differentiating between customer churn and non-churn [73], [74]. RFM-based variables are often included as features in models for customer loyalty, defection, and customer lifetime value prediction [74]. Reference [75] used time-varying RFM using deep learning to predict churning customers in the financial sector. Some studies have expanded upon the RFM features by incorporating additional variables. For instance, the length of the customer relationship was added to the RFM features, resulting in LRFM features [76], [77], [78].

In mutual funds, Maldonado et al. [79] derived RFM-based features from customer behavioural activities such as seniority, redemption, and repurchase actions. In this context, the combination of financial indexes and customer demographic information with RFM-based features was considered. Reference [80] identified five additional variables, in addition to RFM, for predicting customer churn. These additional variables include the number of purchased items, returned items, discount, prize, and distribution time. Similarly, financial variables and indexes were integrated with behavioural and demographic features as churn predictors, including daily

[6]is an undersampling technique that selects similar instances from the majority class to the minority class to improve the model's ability to distinguish between the classes.

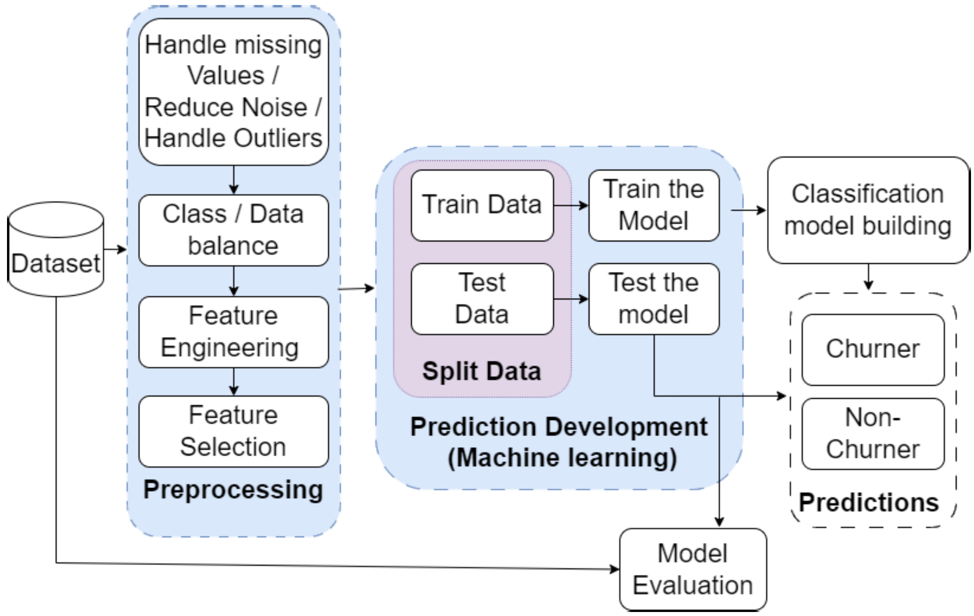


FIGURE 5. **End-to-end churn prediction model architecture based on CRISP [57].**

returns, Standard & Poor's 500, currency exchange rate, and the selective price of shares index (IPSA). Similarly, [81] added diversity of customer purchasing behaviour, extending the RFM features (RFM-D). Additionally, [76] demonstrated that monthly revenue, revenue development in the last month and the days since the last contact are important churn predictors. Chen et al. [82] investigated the importance of adding profit to the LRFM variables, creating LRFMP features for predicting churn in the financial sector. They considered the time between transactions as a time variable (L) and demonstrated that it significantly impacts customer churn. Reference [83] proposed another variation of LRFMP by incorporating periodicity and seasonality in sale interactions. In another study, Mirkovic et al. [77] combined QoS and web platform usage features with the LRFM features. The LRFM features, based on invoice data, were combined with QoS to create additional predictors of customer churn [84]. Similarly, a recursive feature elimination method using SVM and random forest was employed, iteratively removing the least important features through backward elimination. Reference [85] extended the RFM feature by adding user lifetime, intensity, and rewards (RFM-LIR) features to capture player-game interactions in the context of churn prediction in the online gaming industry.

Kaya et al. [46] also highlighted the superior predictive performance of dynamic behavioural features compared to demographic features across various business domains. Their study, focusing on the financial sector, specifically considered dynamic behaviours, such as spatiotemporal patterns of expenditure and transaction modes. Expenditure patterns provide insights into the diversity, loyalty, and regularity of spending habits regarding time and location. At the same time, transaction mode choice captures the distribution of transactions across different channels, including online, credit/debit card, and fund transfer methods. By examining these factors from the perspective of merchants, spending categories, and location, a more comprehensive understanding of customer churn behaviour can be obtained [46]. In [65] and [66], researchers emphasised the selection of temporal features based on multivariate time series as important predictors for churn. Additionally, [65] focused on various behavioural features, such as customer spending and usage patterns within different time windows (daily, weekly, and monthly). In their study, RFM and statistical features were employed to represent each series of multivariate time series, with RFM features exhibiting superior predictive performance compared to statistical features. Moreover, it was found that RFM features based on daily dynamic behaviour outperformed those based on weekly or monthly behaviour. On the contrary, some studies indicate that RFM features may be more important than structural features, with no clear evidence that either leads to better predictive performance [71]. It is suggested that using diverse feature types improves churn prediction performance [86], and some studies have combined RFM-based temporal behavioural interaction features with structural features to improve the predictive capabilities of customer churn models [71]. However, extracting structural network features from call graphs can be both computationally and methodologically expensive.

The RFM-based features, while useful, focuses solely on purchase data within a specific period and does not consider customer-company interactions beyond the sales process, which are crucial for developing customer-centric strategies.

Another key challenge with RFM-based features is that RFM variables are not useful for actionable marketing but only for predicting customer churn [107]. To address this limitation, the Recency, Frequency, Importance, and Duration (RFID) features were proposed [128], [129] as a complementary approach incorporating customer interactions in predicting churn. Hasumoto and Goto [107] extended the RFM features (eRFM) by adding three additional features: store, tenure, and inter-purchase time. They linked churn with customer purchase behaviour and highlighted that these variables are useful for practitioners to take marketing actions. Reference [66] leveraged social network analysis techniques to extract customer relationships by calculating centrality measures in temporal graphs based on structural call graph features. Other studies have explored the use of customer information (demographics and contact details) and customer data usage to develop churn prediction models [15].

Historical purchasing and bank account transaction data have been employed to enhance marketing strategies and retention campaigns [15]. Moreover, unstructured text data has also been utilised to detect customer attrition in financial services [130], [131]. Similarly, [132] utilised unstructured text-based customer feedback data to develop a recommendation system to enhance customers' loyalty to repair services. The authors applied an opinions mining algorithm to extract sentiment scores and transform the data into a structured format. Subsequently, an action rule mining algorithm processed the structured data derived from sentiment mining. Vo et al. [130] extracted four feature sets from customer call log data: term importance, lexical information, phrase embedding, and personality traits. To capture informative terms, they used term frequency and inverse document frequency and filtered out stop-words.[7] However, term importance alone does not provide enough information concerning the customer. Therefore, phrase embedding was applied to capture the semantics of the text. Additionally, Sentiment analysis was also carried out to extract latent concepts and topic-related features, which can identify the text with either positive or negative emotion, and these features are essential to capture customer satisfaction. Finally, the authors employed a five-factor personality model to extract personality traits and associate them with a customer's financial decision for customer churn prediction [130].

De Caigny et al. [131] proposed two different strategies to extract features from unstructured text data: the vector space-based approach and the convolutional neural network (CNN) based approach. They performed text cleansing, tokenisation, part-of-speech tagging, and stemming in the former. Subsequently, they excluded irrelevant and low-frequency terms. In the latter, they applied one-hot encoding to convert textual information into numerical data. Recently, Wang et al. [136] claimed that personality characteristics of users should also be considered when predicting customer churn. They combined images (user avatars) with textual features, taking into account personality traits. They trained a deep-learning model using the Flickr dataset containing sentence-based image descriptions. Findings suggested that both text and image improve the overall performance as measured by precision, recall, and F-score metrics.

Recent studies on customer churn have integrated diverse data types to analyse customer attrition, including demographic, spatio-temporal, and behavioural data. Factors such as customer-company relationships, marketing and financial indicators have been recognised as relevant for distinguishing potential churners from non-churners [40]. In subscription-based companies, the length of the customer-company relationship plays a crucial role, with longer relationships indicating a lower probability of churn [133]. Customers who have renewed their contract at least once are less likely to churn. These findings emphasise the importance of considering customer-company interactions and the longevity of relationships in predicting churn behaviour. Recently, [117] found that service quality, customer satisfaction, subscription plan upgrades offers, and network coverage are the prominent churn drivers in the Danish telecommunication industry.

To summarise, the important features utilised in churn prediction, as identified in the literature, are outlined in Table 3. Demographic and RFM features and their variations emerge as prominent choices across various industry sectors.In particular, the gaming industry strategically utilises activity logs as a readily available feature, offering insights into players' behaviour. This spans the time invested in gaming and specific activities undertaken while playing. Essentially, feature engineering boosts model accuracy in predicting customer churn while managing computational costs [116]. Although studies indicate that more features can improve predictive performance, the relative importance of different features remains a topic of debate [86]. Practitioners must balance accuracy and operational efficiency by carefully selecting features that capture essential information for churn prediction unless computational power is not a priority.

C. FEATURE SELECTION

Feature selection plays a crucial role in the data preprocessing stage by identifying the most relevant and important attributes. Datasets often contain multiple variables that may be irrelevant, such as user ID, and redundant variables that are highly correlated to each other.

Selecting relevant features can also benefit from domain-specific expertise and the task context. The use of using a feature selection method offers numerous advantages. Firstly, reducing the number of redundant features can save computational time and enhance operational efficiency, especially in real-world industrial scenarios. Secondly, it can maximise model performance by minimising error or maximising classification accuracy [43] and helps prevent the model from overfitting[8] [155]. Also, having fewer

[7]particular terms that do not carry important textual context (such as prepositions or articles).

[8]Overfitting occurs when a model becomes too complex, closely fitting the training data but underperforming on new and unseen data [163].

TABLE 3. Summary of feature usage.

Feature Type	Citations	Remarks
Service/data usage and CDR	[8], [31], [36], [41], [42], [56], [58], [65]–[67], [71], [87]–[105]	Service usage and CDR features are primarily utilised in the telecommunication industry. CDR is easily available for telecommunications, and service usage data is easily available in various industries, such as the telecommunication, financial, and online gaming sectors.
RFM & Variations	[31], [40], [45], [46], [65], [71]–[86], [106]–[109]	RFM features provide insights into recent activities, usage frequency, and monetary value. Although excellent indicators, they are not universally available across industries.
Demographic	[5], [7], [8], [12], [31], [40], [45], [46], [54], [62], [75], [79], [89], [91], [94], [97], [103], [105], [108], [110]–[115]	Demographic features, such as age, prove insightful for identifying churners. Churn likelihood is lower in older individuals than younger ones; however, demographic variables alone may not suffice.
QoS	[84], [106], [116], [117]	QoS variables, although uncommon among various industries, can enhance predictability when combined with other features.
Activities, Logs and Time Spending Behaviour	[59], [60], [63]–[65], [68]–[70], [80], [118], [119]	Predominantly used in the online gaming industry, these variables include total time spent in-game, playing frequency, and game logs data.
Financial Variables and Transaction Patterns	[5], [7], [12], [46], [61], [77], [79], [94], [110], [111], [113]–[115], [120]–[126]	Mostly available for the financial industry, these variables encompass mode of payment, statement balance, frequency, geographical location, income, and expenditure patterns.
Structural Call Graph Features	[40], [66], [71], [72], [127]	Extracted from customer interactions, such as CDRs, where customers are nodes and their interactions through calls/messages serve as links. Customers with links to churners have a higher probability of churning.
RFID	[128], [129]	RFID features complement RFM features, incorporating customer importance and interaction duration with the Contact Center.
Customer feedback and communication data	[75], [105], [130]–[135]	Unstructured data in the form of customer feedback, complaints, or communications, enhances predictive performance in customer churn classification.

independent features promotes understanding and supports model interpretability [37] and therefore enhances its explainability [164].

In the financial sector, Mirkovic et al. [77] proposed utilising invoice-level data and applied a permutation-based feature importance approach to select the most important features. Similarly, [42], [149], [150], [152] utilised Principal Component Analysis (PCA) for data dimensionality reduction to retain those features explaining the largest amount of variance. Specifically, PCA generates new orthogonal variables, called principal components, each explaining a portion of the variance in the data. Those with minimal scores are typically removed, effectively reducing the number of features. Similarly, [80], [146] suggested that discriminant analysis is a robust method for selecting important variables. The Fisher score is used to rank independent features, determining their relative importance for selecting a subset [31], [35], [95], [122].

Reference [137] applied the filter and wrapper method to select optimum features with random oversampling. This technique involves randomly selecting instances from the minority class and duplicating them within the dataset to tackle the class imbalance. Essentially, it duplicates data points from the minority class randomly to achieve a more balanced distribution. Adnan et al. [100] applied various transformation methods, including log, box-cox, Z-score, and ranking, for cross-company churn prediction. They demonstrated that these methods, except for Z-score normalisation, can improve data normality and classifier performance in most cases.

Similarly, Sana et al. [153] compared six transformation methods, including log, rank, box-cox, z-score, discretisation, and weight-of-evidence. They applied these methods to train eight different machine learning classifiers. The results showed that using these transformations improved the predictive performance of the classifiers when compared to their counterparts that were trained without any transformation. Also, [147] demonstrated that z-score transformation led to improved predictive performance of churn prediction models.

Numerous studies have employed heuristic and metaheuristic-based optimisation techniques to select the best features for their churn prediction models. For instance, Bano et al. [96] applied chaotic salp swarm optimisation, which incorporates chaotic dynamics into the collective movement of a salp swarm to enhance exploration and exploitation during the optimisation process. Sudharsan and Ganesh [92] used a hybrid of Brownian movement with butterfly optimisation, which simulates the food foraging and finding mating partner behaviours of butterflies. Shourbaji [89] et al. combined ant colony optimisation (ACO) with the reptile search algorithm (RSA) for feature selection in the telecommunications industry. The concept behind ant colony optimisation is rooted in the collective foraging behaviour of ants, where pheromone communication is employed to find optimal paths in a search space. Similarly, [88] applied chaotic pigeon-inspired optimisation for feature selection. The intuition behind chaotic pigeon-inspired optimisation is to introduce chaotic dynamics into the decision-making process of a swarm, mimicking the adaptive behaviour of pigeons, to improve exploration-exploitation trade-offs in optimisation. Also, Vijaya et al. [90] applied particle swarm optimisation, simulated annealing, and a hybrid of particle swarm optimisation and simulated annealing optimisation for

TABLE 4. Summary of feature selection techniques.

Feature Selection	References	Remarks
Feature Importance	[6], [31], [35], [41], [43], [45], [51], [52], [55], [59], [63], [68], [69], [71], [77], [78], [84]–[86], [90], [95], [97], [101], [117], [122], [137]–[148]	Feature importance is a technique to identify the most influential features in decision-making. Methods include Fisher score, chi-square test, recursive feature elimination, sequential forward selection, correlation, and rank features.
Feature Reduction	[31], [34], [36], [42], [59], [69], [80], [95], [98], [115], [125], [146], [147], [149], [150], [150]–[152]	Feature reduction aims to decrease the number of features through transformations like PCA and LDA.
Feature Transformation	[61], [100], [147], [153], [154]	Feature transformation converts features between representations, using methods like log, rank, box-cox, z-score, discretisation, and weight-of-evidence.
Heuristic Optimisation	[51], [88], [89], [92], [96], [154]–[160]	Heuristic optimisation, e.g., Genetic algorithms or particle swarm optimisation, optimises feature selection process based on a fitness function.
Domain Knowledge	[36]	Domain experts choose important features based on their industry knowledge and expertise.
Deep Learning-based Approaches	[65], [87], [102], [119], [143], [161], [162]	Deep learning, especially auto-encoders, can perform automatic feature selection.

feature selection. The hybrid optimisation algorithm combines particle swarm optimisation and simulated annealing to synergistically exploit the strengths of both techniques in search space exploration. The results revealed that the optimised feature selection and random undersampling improved churn predictability across various performance metrics.

Chen et al. [156] combined the Cox model with various variable penalties such as lasso, smoothly clipped absolute deviation, and minimax concave penalty to predict customer churn in vehicle insurance companies using personal information and behaviour data. Jiang [51] used a modified multi-objective atomic orbital search and extreme learning machine for feature selection in churn prediction. The modified multi-objective search is guided by the principles of quantum mechanics, combining atomic orbital-like structures to efficiently explore the multi-objective optimisation space. Amin et al. [93] used a combination of genetic algorithm and Naive Bayes to select the optimal set of features for predicting churn in the telecommunication industry. They compared their approach with several state-of-the-art techniques and found that their method outperformed the others. Similarly, Sadeghi et al. [108] used Principal Component Analysis (PCA) and particle swarm optimisation in combination with K-means algorithm to optimise the initialisation of centroids and reduce the dimensionality of a dataset. This approach aimed at enhancing the clustering performance by finding the best centroids and reducing the feature space. The synthesis of the feature selection techniques for customer churn prediction is presented in Table 4.

IV. PREDICTIVE MODELING FOR CUSTOMER CHURN

After pre-processing a dataset, the next step is to train a data-driven model for customer churn prediction. The effectiveness of this prediction depends on the data quality, the choice of classifier and the hyper-parameters used to train it [31]. Over the years, various modelling techniques have been employed for churn prediction, ranging from simple and interpretable to complex and less interpretable. Table 5 synthesises and compares the relevant modelling techniques for customer churn prediction, while the following sections discuss their strengths and limitations.

A. INTERPRETABLE LEARNING TECHNIQUES AND MODELS

Interpretable models exhibit a higher degree of transparency and explainability, making it easy for humans, for example, to understand the reasons behind their specific predictions or decisions [165]. Among interpretable learning techniques, logistic regression and decision trees are commonly employed for churn prediction due to their straightforward functioning [31]. The weights (coefficients) associated with each independent variable in logistic regression indicate changes contributing to the prediction. In the case of decision trees, exact decision rules can be extracted directly from the tree structure, simplifying the interpretation process. The next sub-sections provide a deeper review of interpretable learning techniques and models for churn prediction.

1) LOGISTIC REGRESSION

Logistic regression is a widely adopted method and a de facto industry standard for churn prediction [40]. Gattermann-Itschert and Thonemann [84] used logistic regression with L2 regularisation for churn prediction in the convenience wholesaler industry. It was trained on multiple slices to avoid overfitting and to make it more generalisable and robust. Multiple time slices were generated for transactional time-stamped data and combined into one training set. A profit-maximising logistic regression was used in [138], which directly integrated the profitability parameters into the model construction. It was based on regularisation with lasso, which maximised profitability during the training process using a real-coded genetic algorithm. This approach yielded optimal performance in profit-maximising, precision, and recall on nine real-life datasets.

Coussement et al. [154] applied logistic regression after data pre-processing to predict churn in the telecommunications industry and compared its predictive performance with more advanced single and ensemble data mining algorithms such as random forest, bagged CART, J4.8 decision tree, multilayer perceptron, Naive Bayes, radial basis kernel SVM and Stochastic gradient boosting classifiers. They tested the impact of fine-tuning parameters on the predictive performance of logistic regression in the case of customer abandonment. They found that proper data preparation and fine-tuning parameters significantly improve predictive performance, making it competitive with advanced models. They argued that developing an advanced model for churn classification is not essential compared to an optimised simple model, particularly with optimal data pre-processing and preparation.

Kim et al. [63] compared logistic regression with ensembles[9] and deep learning.[10] They concluded that there is little relationship between the choice of predictive model and prediction performance. In the game-based learning dataset [166], logistic regression outperformed the random forest classifier. Similarly, logistic regression showed competitive performance with LightGBM and performed significantly better than random forest on a financial dataset [167]. [109] applied logistic regression with a mixed penalty to avoid overfitting, and their results showed that adding a mixed penalty to logistic regression improves its performance over simple logistic regression. Uner [103] applied binomial logistic regression analysis to predict churn based on discrete choice theory. Their objective was to find churning factors, and their finding showed that customer service factors such as call quality, billing and brand image affect both loyalty and churn intentions.

2) DECISION TREES

Decision trees are another popular predictive learning technique in business analytics due to their high interpretability by design [168] and robust rule extraction method. They have been applied, for example, in future stock market prediction and customer-related decision-making [169]. The decision tree technique iteratively splits the instance space into smaller subgroups by applying optimal criteria to classify data into different classes, such as churners and non-churners. It begins with a root parent node and recursively splits the data into child nodes based on split criteria such as entropy, impurity or information gain. This process continues until no further splits are necessary, terminating at a leaf node. Each node in a decision tree is represented by an attribute and a condition for the split. Popular decision tree-based learning techniques for churn prediction include the C4.5 and C5.0 [143], [170]. These algorithms split instances based on information and entropy, creating smaller subgroups. The tree potentially grows to its full length before pruning to enhance generalizability on unseen instances.

[9] gradient boosting, random forest.

[10] CNN and long short-term memory (LSTM).

Another variant of the decision trees is the Classification and Regression Tree (CART). CART recursively splits instances and records the impurity of the parent against the child node to determine the goodness of the split (low impurity indicates a good split). A parameter controls the threshold for impurity, which determines the size of the tree until the impurity is lower than the threshold. This technique has been used, for example, by [144] to predict churners in the telecommunications industry. Similarly, Al-Najjar et al. [143] compared C5.0 with Artificial Neural Networks (ANNs), Bayesian trees, Chi-square Automatic Interaction Detection (CHAID) trees, and Classification and Regression (CR) tree, demonstrating that C5.0 exhibited the best predictive performance on a financial dataset. Alizadeh et al. [5] applied five different decision tree variations (C4.5, C5.0, CART, CHAID, and CTree) to extract churn behavioural rules from a bank customer dataset. CTree yielded the best results based on accuracy, F-measure, and ROC evaluation criteria. After extracting rules from the decision tree, the behavioural patterns of selected customers with churner behavioural rules were compared using the change mining method and its similarity measure. The method is used to analyse and extract patterns or insights from changes in data over time. The opinions of banking experts were also incorporated and combined with the model's predictions to obtain a final churn score.

Although decision trees are the most common choice for churn prediction [171], they cannot fully capture complex nonlinear relationships among the features of a dataset [170].

3) NAIVE BAYES

Naive Bayes is a class of probabilistic classifiers based on statistical learning based on Bayes' theorem, assuming independence among variables. Naive Bayes is a constrained form of a Bayesian network, assuming that predictor variables make an independent and equal contribution to the outcome. Amin et al. [172] used a Naive Bayes classifier to analyse customer churn behaviour in the telecommunication industry. They partitioned a dataset into upper and lower distance*zones* based on distance values. Their findings showed that the classifier exhibited high decision certainty and accuracy on the data with greater distance.

Ullah et al. [101] analysed customer churn in the telecommunication industry and compared the performance of Naive Bayes with various other classifiers. Their findings suggested that Naive Bayes was the worst-performing among all classifiers. Similar results were observed in [173], where Naive Bayes performed worst compared to eleven other classifiers. On the contrary, Amin et al. [100] reported Naive Bayes as the best-performing compared to the other employed classifiers, including k-nearest neighbours (KNN), GBT, single rule induction and deep learning for cross-company churn prediction in the telecommunication sector. They compared all models with and without data transformation using log, boxcox, Z-score and rank-based techniques.

B. NON-INTERPRETABLE MODELS

Non-interpretable models are complex models that lack transparency in their decision-making mechanism. These models are also known as black-box models because they are not inherently interpretable, their decision-making procedure is not understandable to humans, and their predictions are not traceable or inferable by humans [165].

1) SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are a set of learning methods used for regression tasks, classification, and outlier detection [174]. SVM separates two classes by drawing a hyperplane called a decision boundary. It uses a kernel trick to transform data and finds a decision boundary to segregate and split it into different classes. SVMs are popular and widely adopted due to fewer control parameters and the ability to generalise the learning process [141]. In general, SVM-based models have good predictive performance, and they are promising for identifying customer churn because they have been shown to perform better than Naive Bayes, logistic regression, and decision trees [170], [175]. Vafeiadis et al. [170] compared ANN, decision tree, logistic regression, Naive Bayes and SVM with and without boosting, the results revealed that SVM and its boosted version could perform even better than ANNs. Kurtcan and Ozcan [42] tuned the hyper-parameters of SVM using grey wolf optimisation. The resulting model was used to predict customer churn in the telecommunication industry and compared with several other classifiers, including logistic regression, Naive Bayes, decision trees, and vanilla SVM. The results demonstrated that the optimised SVM significantly outperforms the other classifiers. However, these results are inconsistent with those presented by [80] and [176], where findings show that tuned ANNs perform better than SVMs [80], and Naïve Bayes outperformed SVM model in terms of precision, recall, and F-score [176]. Additionally, SVMs tend to perform poorly on overlap classes. The SVM's non-linear variant (kernel) is often preferred but lacks full interpretability [38]. The reasons for this poor performance include the reliance on linear decision boundaries, sensitivity to outliers, and the challenge of capturing the complex non-linear relationships needed to accurately separate overlapping classes.

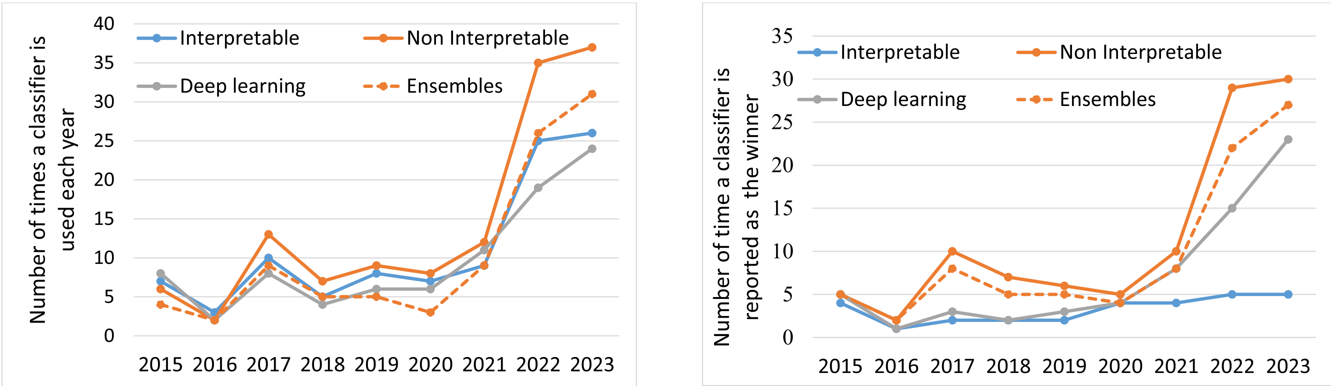
2) ENSEMBLE METHODS

Ensembles have shown to be effective in improving the predictive performance of customer churn. Ensemble learning is a general meta approach that combines the predictions of multiple models to produce better results. Ensembles can be categorised as homogeneous or heterogeneous. The former involves combining the same type of base models, trained on different subsets of data or feature sets, such as bagging and boosting or rotations ensembles. The latter combines different base classifiers such as decision tree, SVM and ANN, but each is trained on the same data or feature sets. Bagging and boosting-based ensembles have been widely used for predicting customer churn [91], [157]. Tree-based homogeneous ensembles are also commonly used for customer churn prediction, including gradient-boosted trees (GBT), random forest, Xgboost, and Adaboost, as shown in Table 5. In this class of learning algorithms [165], an ensemble of weak learners is applied, with decision trees being a popular choice. Formally, a weak learner is a classifier that achieves slightly better than 50% accuracy. In random forest, multiple decision trees are combined through voting, while boosted variants, such as GBT, AdaBoost, and Xgboost, sequentially correct the errors of previous trees. These strategies help minimise the risk of overfitting and enhance predictive capacity.

In the context of mobile social games [118] and the telecommunication sector [118], GBT has been found to outperform other classifiers, which is consistent with claims made in a different study [119] that GBT is popular among ensembles due to its high accuracy and efficiency. However, other studies [69], [76] found that random forest outperformed seven other classifiers, including GBT, in the telecommunication sector. Wahul et al. [177] compared various ensemble methods and found that GBT had better predictive performance than Adaboost, random forest and stochastic gradient descent. Similarly, the random forest technique outperformed decision tree C5.0 in a telecommunication dataset [178], and logistic regression, SVM, and xgboost on a banking dataset [179]. Also, [105] recommends using the random forest for churn prediction in the telecommunication industry. In [180], authors developed an ensemble of random forest and Adaboost for early churn prediction in the telecommunication sector and compared the performance with several other classifiers. Their results revealed that RF-Adaboost outperforms all the other classifiers on several evaluation metrics. Moreover, [68] compared ANN and random forest models for classification and regression to predict churn in similar industries using in-game activity data and showed that random forest outperformed ANN.

In the financial sector, Lemos et al. [120] compared the performance of logistic regression, KNN, SVM, elastic net, decision tree, and random forest and reported that the latter yielded the best results across various metrics. In the same industry, random forest was reported as the best classifier, outperforming other classifiers such as logistic regression and SVM [76], [77], [94]. In a study by [94], which compared over a hundred classifiers to predict customer churn in the telecommunication industry, it was observed that regularised random forest and bagging random forest achieved the highest accuracy and area under the receiver operating characteristics (AUC), respectively, and outperformed other classifiers.

The effectiveness of tree-based ensembles for churn prediction in the telecommunications sector has also been demonstrated in studies such as [45], [46], and [65]. Also [181] compared the performances of five tree-based ensembles, including Adaboost, GBT, xgboost, CatBoost and lightGMB, and found that xgboost outperformed the



(a) Classifiers Applied

(b) Winner Classifiers

FIGURE 6. Classifiers used and Winners between 2015 - 2023 (2024 is not included here as the year has just started. However, a similar trend is observed with the prevalent use of ensembles and deep learning methods).

other classifiers. Similar results were reported in [127], [182], [183], and [184], where Xgboost outperformed logistic regression, SVM, random forest, KNN and other tree-based techniques. Similarly, stochastic gradient boost (SGBoost) outperformed KNN, logistic regression, and random forest on a telecommunication dataset [185]. In contrast, [186] applied xgboost to a telecommunication churn dataset, achieving 99% accuracy.

Hason et al. [145] employed ensemble ANN with the Bayesian network and compared their results with several other tree-based classifiers, concluding that random forest and ensemble classifiers outperform the others. Also, [113] proposed an ensemble by combining multiple bagging and boosting classifiers to enhance the accuracy of churn prediction without overfitting the model. The model selected the three best-performing classifiers among regression, decision tree, SVM, and ANN in the bagging step. In the boosting step, an associative classifier with an apriori algorithm was applied to increase the accuracy further. During testing, wrongly classified instances from the bagging classifiers were fed into the booster classifier using associative rules.

Similar findings were observed in [91], where authors compared different bagging and boosting ensembles with various other classifiers, demonstrating their superiority. Similar results were also observed in [125], showing that boosting-based ensembles, specifically Light Gradient Boosting Machine (LightGMB), outperformed random forest, decision tree, SVM, and logistic regression on multiple financial (banking) and telecommunication sectors datasets.

Bogaert et al. [31] compared the performances of several standalone, homogeneous, and heterogeneous classifiers using various evaluation metrics considering churn datasets from different sectors. The results revealed that the heterogeneous combination had better predictive performance and gradient-boosted models were highly ranked among homogeneous ensembles in line with the finding observed in [118]. Likewise, similar results were observed in [121] where the light gradient boosting machine (GBM) outperformed the SVM, random forest and xgboost on a financial dataset. Furthermore, the authors also improved light GBM by adding a focal loss function to handle the class imbalance problem that focuses on minority classes and difficult-to-score samples. The loss function tries to balance positive and negative classes to overcome the classification problem of unbalanced data. AlShourbaji et al. [160] proposed an ensemble of SVM and GBM, which uses SVM as a base learner and decision trees as weak learners in the GBM's structure. The model's hyperparameters were optimised using a modified version of particle swarm optimisation, and the results were compared with SVM and GBM on seven churn datasets. The findings indicated that the proposed model exhibited significantly better predictive performance compared to other classifiers.

In [112], the predictive performance of standalone and ensemble classifiers was compared on telecommunication datasets. Boosting tree ensembles outperformed standalone classifiers, leading to the proposal of a two-layer new classifier. The first layer consists of two blocks: the first comprises four classifiers (KNN, Naive Bayes, logistic regression, and RF), and the second includes four ensembles (Xgboost, cat boost, GBT, and extra tree). Soft voting is performed within each block, and the outputs are further combined in second block using another soft voting. Findings demonstrated that the ensemble of classifiers outperformed the individual classifiers.

De Caigny et al. [35] proposed a hybrid[11] predictive technique called the logit leaf model and tested it on fourteen different datasets from financial, telecommunication, energy, newspaper, and DIY supplier services. The primary idea behind this approach was to train various models on different data segments instead of the entire dataset, assuming that it could result in better predictive performance. This hybrid learning technique works in two phases: segmentation and prediction. In the first stage, the technique creates decision rules for customer segmentation that identify homogeneous

[11]Ensemble methods work independently to vote on an outcome while hybrid methods work together for prediction.

customer segments. In the second stage, logistic regression is applied with a forward variable selection method for each segment at the terminal node. The decision to develop the logit leaf technique stemmed from the interpretability and predictive performance offered by both the decision tree and the logistic regression. This hybrid technique was shown to have better predictive performance than its baselines (logistic regression and decision tree) and is at least as good as advanced ensemble methods such as random forest and logistic trees [35].

In summary, ensembles are generally more efficient than simpler models but often lack interpretability [187], [188]. To mitigate this, post-hoc or ante-hoc methods are often employed [129]. There are limited works dedicated to the interpretability or explainability of ensembles such as Spline-rule ensemble [38], logit-leaf model [35] and SHapley Additive explanations (SHAP) [125], [130], [148], [189] in the context of churn prediction.

C. DEEP LEARNING

Several recent studies have used deep learning techniques to analyse customer churn [110], [123], [145], [190], [191], [192]. Despite their effectiveness in training highly accurate models, these techniques lack interpretability due to the vast number of parameters and the long chains of non-linear transformations involved. This complexity makes the compression of how input data is mapped to the output difficult.

Khodabandehlou et al. [80] demonstrated that deep learning outperformed SVM and decision tree C5.0 in predicting customer attrition on a food industry dataset. In the financial sector, deep learning outperformed random forest and SVM, showing significant performance improvement of 11% and 15% respectively [123]. Similarly, [193] demonstrated that deep learning outperformed various other standalone and ensemble classifiers on telecommunication datasets. Also, [161] proposed hybrid probabilistic possibilistic fuzzy C-means clustering (PPFCM) along with ANN to anticipate customer abandonment in telecommunication industry. In the training phase, PPFCM was used to cluster the training data into clusters known as the cluster training set. Then, separate ANNs were trained on each cluster for the classification. Following the training process, the test data was also clustered similarly into a cluster test set. In the test phase, the corresponding nearest ANN classifier was selected based on the maximum similarity between the cluster training set and the cluster test set. The performance of PPFCM clustering was compared with various other classifiers, including K-means clustering, fuzzy C-mean, Possibilistic C-means and Possibilistic fuzzy C-means clustering techniques, resulting in the highest with the lowest error. Then, the ANN was trained using the Levenberg–Marquardt (LM) algorithm, and the performance was compared against other standalone and hybrid classifiers, demonstrating it was the highest.

Ahmed et al. used a pre-trained CNN model from ImageNet data to classify telecommunication datasets [102]. To do this, they transformed one-dimensional feature vectors into two-dimensional images. In these images, each row represented a day, and each column represented a specific type of behaviour, such as time and duration of calls, voice mails, international calls, and recharge. The authors fine-tuned the pre-trained model using transfer learning to improve classification accuracy. Also, in the financial sector, Wangperawong et al. transformed time-varying features into images and then applied CNN to effectively predict customer churn [162]. Khodadadi et al. [124] used Recurrent Neural Networks (RNN) for churn prediction in advertisement sector, and [75] applied different architectures of deep learning, namely RNNs and Transformers, using time-varying RFM features, and reported dominance of RNNs. Similarly, Ramesh et al. [142] found that ANN performed better than tree-based models, Naive Bayes and KNN. Additionally, [194] proposed a hybrid deep learning architecture by combining CNN and Bidirectional LSTM. They compared its performance with Bidirectional LSTM, RNN, KNN, decision tree, SVM, and bagging- and boosting-based ensembles. The findings revealed that the hybrid model outperformed all the other models, suggesting that deep learning can significantly improve the development of predictive models with higher accuracy compared to traditional machine learning methods.

Besides various deep learning architectures, optimisation techniques have played a crucial role in improving the performance of deep learning models. Mirza et al. [87] used a deep canonically correlated autoencoder for churn prediction and optimised its hyper-parameters using deer hunting optimisation. The intuition behind deer hunting optimisation involves mimicking the foraging behaviour of deer to iteratively optimise solutions in search spaces through a combination of exploration and exploitation strategies.

Similarly, [195] optimised parameter tuning of a deep neural network using the Bat optimisation, which draws inspiration from the echolocation behaviour of bats where virtual bats explore and exploit a search space using echolocation-inspired movements. Also, [140] optimised the weight of the deep convolutional neural network using a hybrid of firefly-spider optimisation and spider monkey optimisation, with recursive feature elimination for feature selection. The intuition behind the hybrid optimisation is that it combines the light-attracting dynamics of fireflies and the intelligent foraging behaviour of spider monkeys, aiming to create a synergistic optimisation approach that leverages their strengths for improved performance in search spaces. Similarly, [159] applied elephant herding optimisation to optimise the parameter tuning of RNN for telecommunication churn prediction. The idea behind elephant herding optimisation is that it harnesses the collective intelligence and cooperative behaviours of elephant herds through collaborative exploration and exploitation of the search space. In another study, [158] optimised the parameter of stacked bidirectional LSTM using an improved gravitational search optimisation algorithm. The optimiser mimics the gravitational forces between masses in physics, using attractive and repulsive forces to simulate the movement of solutions in the search space, aiming for optimal solution convergence. The authors [158] compared the result of their optimised RNN with CNN and simple RNN, showing that optimised RNN significantly outperforms CNN and RNN on all three metrics: precision, recall, and F-measure.

Rodan et al. [196] proposed a neural network ensemble with negative correlation learning in the telecommunications sector. They compared the performance of their approach with fifteen other classifiers and showed superior performance of the proposed ensemble on three metrics: prediction accuracy, hit rate, and actual churn rate. Moreover, the result also showed that AdaBoost, SVM, and bagging performed competitively with the ensemble methods, whereas Naive Bayes showed the worst performance. Similarly, in [134], transformers and their various variations were compared with SVM, logistic regression, Naive Bayes, and multilayer perceptron on a security churn dataset, revealing that transformers outperform the other classifiers. Also, [152] applied deep feedforward neural networks with the Block Jacobi SVD algorithm for customer churn prediction in a cloud environment,[12] demonstrating that deep learning outperforms decision tree, logistic regression, random forest, xgboost, and Naive Bayes on accuracy, recall, precision, and recall.

Although deep learning has shown promise in developing models for churn prediction, they often lack interpretability, making it difficult to understand their inferential process. The lack of transparency to provide direct insights to support human understanding also limits their adoption for churn prediction [189]. As a result, recent research has focused on mitigating these issues. Liu et al. [197] have developed an interpretable deep learning model for churn prediction in mobile social applications. They used LSTM, a variant of RNN that processes long-term dependencies in time series data. An attention-based mechanism highlighted the important features that most contributed to a particular prediction. Other methods, such as partial dependency diagrams [129], [198], feature Importance [129], Local Interpretable Model-agnostic Explanations (LIME) [129], [199], and SHAP [126], [189], [199] have also been proposed to enhance interpretability and support the explanation of deep learning models for churn prediction.

Recent years have seen an abundance of strategies for churn prediction, including various machine learning classifiers, survival analysis, regression analysis, and statistical approaches [139]. Besides this, factors such as the selection of features, the underlying modelling approach, and the specific data used for training can all lead to significantly different outcomes. Moreover, pre-processing and sampling techniques [68] and the definition of churn itself [85] have notable impacts on model performance. To be precise, with so many options in the end-to-end prediction model architecture (see Figure 5, selecting an appropriate model can be challenging [200].

To tackle this challenge, it is essential to stay up-to-date on the latest trends. Figures 6a and 6b demonstrate the annual trend of baseline classifiers and winning models reported in the literature from 2015 to 2023, respectively. The trend highlights that interpretable, non-interpretable and deep learning-based approaches have been increasingly adopted in recent years. Notably, most churn prediction models were developed using deep learning, which are non-interpretable models. Also, among non-interpretable models, homogeneous and heterogeneous ensembles are consistently adopted more than other standalone non-interpretable models.

Predictive performance of deep learning techniques has shown improvement from 2015 to 2023, with a slower trend from 2015 to 2020, followed by a sharp improvement thereafter. In 2023, the trend suggests a growing preference for deep learning techniques, which, when incorporated among the baseline models, are more likely to emerge as winners than non-interpretable models, including ensembles. Reported winners are typically determined based on evaluation metrics such as accuracy, F-score, LIFT, expected maximum profit, and AUC. A summary of relevant literature reviews, including objectives, methodology, feature selection, sampling strategies, and remarks, is provided in Table 6.

V. MODEL VALIDATION

After building a prediction model, the next step is to evaluate its predictive performance. Model validation is the process of estimating the effectiveness of a model's prediction ability over unseen data. A straightforward principle of model validation is to split the data and hold some data for validation. The hold data that is not used during training is used to evaluate the model's predictive performance. The model's efficacy can be measured based on various matrices such as accuracy, precision and recall. In this section, we will discuss different ideas and techniques related to model validation.

A. CROSS-VALIDATION AND TEST DATA

1) CROSS-VALIDATION

Cross-validation (CV) is a commonly used sampling method to evaluate the generalisability of a model over a dataset [15]. Its core principle is to use all available data to train and test the model. Examples of CV techniques are K-fold CV and Monte Carlo CV.

- *K-fold cross-validation* - In this technique, a dataset is randomly divided into K subsets of equal size. Each subset is held out once as a validation set, while the rest of the data is used to train the model. For example, Moeyersoms et al. [33], [69] used 10-fold CV, whereas Höppner et . [37] performed five replications of 2-fold CV (5×2) on the training set, with each

[12] Also, the model was hosted on a cloud environment, allowing users to access the model from anywhere with an internet connection.

fold stratified according to the target feature. Leave-one-out CV was also used in [151]. Leave-one-out CV is an extreme version of a K-fold CV, where a model is trained and evaluated for each single instance in the dataset. However, this variation of K-fold CV carries maximum computational complexity. While K-fold cross-validation is known for its low variance but higher bias, it aids in mitigating overfitting by dividing the data into distinct subsets. Despite its advantages, K-fold CV has some limitations, including its computational demands stemming from training and testing the model multiple times. Additionally, it might not be suitable for certain types of data, such as time series or spatial data, where the order or location of observations plays a crucial role. If the data distribution across folds is uneven, it could introduce bias into the evaluation process.

- *Monte Carlo cross-validation* - In this technique, a dataset is randomly split into two subsets, with one used for training a model and the other for its validation. This splitting process is repeated many times independently. Vafeiadis et al. [170] used the Monte Carlo cross-validation sampling strategy in the churn prediction problem. Unlike K-fold CV, where each instance in the dataset is tested only once, and the number of partitions is limited by K, Monte Carlo CV tests each instance arbitrarily, so a large number of partitions are possible. This method has low variance but high bias, and the limitation of this method is that a particular instance could be left while some instances could be repeated.

2) TEST DATASET

It's a common practice to split the data into 80% training and 20% test data [201]. In data science, a typical split is two-thirds of the data for training and one-third for test [202]. Several authors in the literature, such as Höppner et al. [37], [77], [120], have successfully employed the test dataset to obtain out-of-sample performance estimates (holdout sample or test data) and validate the effectiveness of their churn prediction models.

B. EVALUATION METRICS

Once the dataset is split into training and validation sets, the model's performance is evaluated on the validation[13] set using different metrics that assess the predictive accuracy of the churn prediction model, as outlined below.

1) ACCURACY

Accuracy is the most common evaluation criterion [45], [203], [204]. It measures the percentage of correctly predicted instances, which means it is the ratio of accurate predictions to the total number of predictions.

$$Accuracy = \frac{Accurate\ predictions}{Total\ number\ of\ predictions} \tag{1}$$

While widely accepted, accuracy does not take into account the class membership probabilities[14] of the predicted class [98]. Moreover, it is not a preferred choice in the case of class imbalance [196] between churners and non-churners, which is very common in the task of churn prediction [95]. Accuracy can be biased towards the majority class on highly imbalanced datasets. For instance, if a dataset has 90% non-churner and 10% churners, the model can achieve 90% accuracy by predicting every instance as a non-churner.

2) PRECISION

Precision is a metric that operates over a single class. In churn prediction, it is commonly used to measure the ratio of correctly predicted churns (true positives) to the total number of predictions classified as churn (true positives + false positives) [37], [45], [59].

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \tag{2}$$

3) SENSITIVITY

Sensitivity, also called recall, measures a model's ability to predict the presence of overall churns in the dataset. It is the ratio of correctly predicted churns to the total number of churns available in the dataset. In other words, the churns predicted as churns are the true positives, while churns predicted as non-churns are the false negatives. Several studies [11], [45], [79], [101], [172] have used recall to compare the model's ability to detect a maximum number of churners over non-churns. This is because detecting churners to the maximum number is considered a major concern for businesses, even if the classifier wrongly predicts some non-churners as potential churners.

$$Sensitivity = \frac{True\ positive}{True\ positive + False\ negative} \tag{3}$$

4) SPECIFICITY

Specificity measures a model's ability to predict the presence of overall non-churns in a dataset. It is the ratio of correctly predicted non-churns, which are true non-churns predicted as non-churns (the true negatives) to the total number of non-churns, that means the true non-churns predicted as non-churns (the true negative and the false positives.

$$Specificity = \frac{True\ negative}{True\ negative + False\ positive} \tag{4}$$

Sensitivity and specificity are often inversely related; if sensitivity increases, specificity decreases and vice versa. Sensitivity denotes how many churners are correctly identified as churners, whereas specificity denotes how many

[13] test set or the crossfold held out set.

[14] i.e., the predicted confidence score obtained from the classifier for an instance belonging to a specific class.

non-churners are correctly identified as non-churners by the model. Higher sensitivity means that there are fewer false negative results. On the other hand, higher specificity indicates fewer false positives.

5) 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 [30], [35], [36], [65], [71], [84], [100], [122]. These metrics provide an objective and systematic way to assess classifier performance without requiring a specific classification threshold value, as multiple values are tested [37]. It is a function of sensitivity to the inverse of specificity.

$$ROC = \frac{Sensitivity}{1 - Specificity} \tag{5}$$

A ROC curve is plotted as a function of sensitivity on the y-axis and the inverse of specificity on the x-axis. A value of 1 indicates that the model can effectively distinguish between churn and non-churn customers, whereas deviations from this indicate a decline in performance. Unlike accuracy, ROC is more suitable for imbalanced datasets, as the AUC measures the classifier's ability to distinguish between churn and non-churn classes based on membership probabilities [205]. The AUC summarises the overall performance of a classifier for all possible cutoffs, making it a widely used metric in churn prediction. However, when the data is severely imbalanced, ROC is biased towards the majority class, which is a common issue in churn data [95], [206]. Therefore, it is important to consider other evaluation metrics in conjunction with ROC/AUC to ensure the effectiveness of churn prediction models.

6) LIFT CHART AND LIFT INDEX

Lift is a crucial concept in churn analysis that helps categorise customers based on their likelihood to churn. This places customers with higher probabilities of churning in the top segment [35]. The lift of a target group reflects the higher proportion of customers who actually drop the service when compared to the whole population of customers. For instance, if 2% percent of customers drop the service, and within the group identified as "churners", 8% drop the service, then the lift is 4.

Lift is a commonly used metric in churn studies, especially in marketing [15], [31], [95], [207], as it helps allocate marketing budgets proportionately to the customers who are more likely to churn [2]. Lift has two alternative variations: the top decile lift and the lift index. The top decile lift segments the highest-risk customers based on the predicted churn probability and arranges the records into deciles according to their prediction score to churn. The top-decile lift metric captures how the occurrence among the top 10% of customers with the highest model predictions contrasts with the overall sample occurrence. Let us assume a company is interested in the top 10% of customers who are at the highest risk of churn based on churn scores. The top decile lift then equals the ratio of churners in the top decile ($D_{10\%}$) ordered by churn score to the total churn rate in the entire dataset (DS).

$$Top - decile\ Lift = \frac{D_{10\%}}{DS} \tag{6}$$

When the lift value is higher than 1, it shows a higher density of churners in the top decile. On the other hand, the lift index is the weighted index of the correctly predicted churners, ranked by their posterior churn probability. Let S be a ranked list of customers based on churn score, then the lift index is calculated as:

$$Lift\ Index = \frac{1.0 \times S_1 + 0.9 \times S_2 + \ldots\ldots + 0.1 \times S_{10}}{\sum_{i=1}^{10} S_i} \tag{7}$$

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 $S_1 = \sum_{i=1}^{10} S_i$ 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 [95], 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.

7) PROFIT-MAXIMISING METRICS

For a business, profit is the ultimate goal. To achieve this, churn prediction strategies need to adapt by applying profit-based metrics instead of conventional classification and evaluation methods [208]. While the accuracy of predictive models is an essential factor, businesses need to approach churn models from a financial point of view.

Profit-based performance metrics were first proposed by Verbeke et al. [209] to measure the effectiveness of a classifier by considering the profitability of a customer retention campaign. Specifically, they evaluate a classifier's performance while considering the cost and expected returns on a retention campaign. Profit-based metrics target a fraction of customers α with a high churn risk. These targeted customers are contacted, incurring an individual cost of c, and are offered monetary incentives. The cost of monetary incentives is d only incurred when the offer is accepted. Only a fraction of customers γ in "true would-be churners"[15] (i.e. with high lift) will accept the offer and stay, whereas all the "false would-be churners"[16] will accept the offer since they had no intention to churn [209].

The benefit of retaining a customer is their customer lifetime value (CLV), which is usually higher than the cost of monetary incentives (d) and the cost of contact (c). Sending

[15] correctly classified as churners, or true positives.

[16] incorrectly classified as churners, or false positives.

TABLE 5. Algorithmic comparison of relevant studies for customer churn prediction (*BT = Boosting trees, DT = Decision tree, LR = Logistic regression, RF = Random Forest, NB = Naive Bayes).

Ref.	LR	DT	BT	RF	NB	KNN	SVM	ANN	Any other	Reported winner	Algorithmic Remarks
[37]		✓							ProfTree, CART, CTree, EvTree	ProfTree	Evolutionary algorithm was used for optimal split on profit-based metrics
[36]	✓	✓	✓				✓	✓	FNN, VQNN, FRNN, OWANN	FRNN	The fuzzy classifiers gains performance on noisy data
[40]	✓			✓				✓	Relational classifiers (WVRN, CDRN, NLB, SPA-RC)	Non-relational classifiers	Relational learners are less flexible than non-relational classifiers
[71]									node2vec representation learning	N/A	node2vec consumes RFM-augmented call graphs to decreases computational complexity.
[45]	✓	✓	✓	✓	✓			✓	Bayesian regression analysis	Adaboost, RF	Ensembles perform better than ANN on small datasets.
[46]				✓					N/A	N/A	RF was trained on spatio-temporal and customer's choice behaviour.
[65]	✓			✓				✓	LSTM and CNN	Deep learning, RF	Raw time series features and RFM-based time series features gain performance for DL and RF, respectively.
[66]									Similarity forests	N/A	Extracted graph structural relationships for training the similarity forests.
[35]	✓	✓		✓					Logistic tree model, Logit Leaf Model	Logit Leaf Model	Trained on data segments instead of the entire dataset to gain predictive performance while maintaining interpretability.
[113]	✓	✓	✓	✓			✓	✓	Hybrid	Hybrid	Eliminate overfitting and dispersion problem by feeding wrongly classified instances from the bagging phase to the association rules.
[91]		✓	✓		✓	✓	✓	✓	Ensembles (BaBaG, BoBaG, BN-NGA)	BaBaG	A hybrid combination of bagging and boosting ensembles improves the predictive performance.
[120]	✓	✓		✓		✓	✓		Elastic net	RF	The algorithm's performance is restricted to a particular dataset.
[96]									Fuzzy Rule-based Classifier (FRC)	N/A	Utilised Quantum-based PSO to select membership functions of FRC model, which improves its predictive performance
[92]								✓	Swish RNN	Swish RNN	The methods are effective in handling diverse uncertainties.
[5]		✓							C4.5, C5.0, CTree, CHAID, CART	CTREE	The performance differences between DT variations are minimal.
[42]	✓	✓			✓	✓	✓		SVM with PCA, SVM with PCA and grey wolf optimisation	SVM with PCA and grey wolf optimisation.	Feature selection and parameter optimisation enhances SVM's predictive performance; however, no interpretability.
[31]	✓	✓	✓	✓	✓	✓	✓	✓	Heterogeneous ensembles	Hetero-geneous ensembles	Heterogeneous ensembles promote high accuracy and diversity by selecting optimal parameters; however, non-interpretable.
[108]									K-means, Ensemble (K-Means with PSO)	Ensemble	Used PSO to optimise the K-Means algorithm in the ensemble to provide initial centroids.
[87]	✓	✓	✓			✓			Ensemble (Deep Canonically Correlated Auto-encoders)	Ensemble	Used deer hunting optimisation to select hyperparameters of the ensemble; however, non-interpretable.
[74]	✓		✓	✓						RF	All classifiers perform equally well. RF was selected as a benchmark for multi-time slicing.
[59]	✓	✓	✓	✓	✓			✓	Various tree-based models, deep learning, linear regression	LSTM and Tree-based models	Improved LSTM classification with feature subset selection, while achieving optimal results through censoring with conditional inference ensembles; however, not interpretable.
[77]	✓	✓				✓				RF	Multi-slicing improves predictive performance.
[69]	✓	✓	✓	✓	✓	✓	✓	✓		RF	Only DT and neural networks exhibited excellent performance, while other ML algorithms performed considerably worse.

TABLE 5. *(Continued.)* **Algorithmic comparison of relevant studies for customer churn prediction (*BT = Boosting trees, DT = Decision tree, LR = Logistic regression, RF = Random Forest, NB = Naive Bayes).**

[125]	✓	✓	✓	✓			✓		Light-GBM	Light-GBM	Used Bayesian and genetic algorithm to optimise hyperparameters of Light-GBM
[112]	✓		✓	✓	✓	✓			Hybrid (Combination of multiple classifiers)	Hybrid	The hybrid classifier combines the advantageous characteristics of multiple techniques.
[122]	✓	✓			✓	✓	✓		ProfTree, ProfLogit, MPM	MPM	MPM is a robust optimisation approach that minimises the worst-case probability of misclassification.
[95]		✓					✓		N/A	N/A	Compared sampling methods for class imbalance problem.
[193]	✓	✓	✓	✓	✓		✓		Extra trees, XGB, GBM, CNN	ANN and CNN	ANN and CNN perform better on large datasets with several features; however, no interpretbility.
[93]	✓	✓	✓	✓	✓	✓		✓	Adaptive algorithm	Adaptive algorithm	Genetic algorithm for optimum feature weighting.
[97]	✓	✓	✓	✓			✓	✓	Ensemble (Top-3, Top-5)	Ensemble	Ensemble of top performing classifiers; however, not interpretable.
[90]		✓		✓	✓	✓	✓		Metaheuristic classifiers	Metaheuris-tic classifiers	Metaheuristic classifier combined with feature selection has better performance.
[121]			✓	✓			✓		FocalLoss LightGMB	FocalLoss Light-GBM	Both light-GMB and FocalLoss Light-GBM performed equally well.
[63]	✓		✓	✓				✓	LSTM, CNN	No big difference	The choice of classifier had little impact on prediction performance.

incentives to non-churners has no benefit, nor contacting the fraction $1-\gamma$ "true would-be churners" who did not accept the offer. The total profit of the retention campaign can be expressed as:

$$P_c(t;\gamma,CLV,\varphi,\phi) = CLV(\gamma(1-\varphi)-\phi)\lambda_0F_0(t) - CLV(\varphi+\phi)\lambda_1F_1(t) \quad (8)$$

Here, $\varphi = \frac{d}{CLV}$ and $\phi = \frac{c}{CLV}$. P_c is the profit of the retention campaign, t is the classification threshold, and γ is the probability of targeted "true would-be churners" who accept the offer. λ_1 and λ_0 are the prior probabilities of the customer as churner and non-churner, respectively. $F_1(t)$ and $F_0(t)$ are the cumulative distribution functions for churners and non-churners. The following are the two profit-maximising criteria:

- *Maximum Profit Criterion (MPC)* - It can provide the maximum profit for a retention campaign for a churn prediction model by selecting the optimal target group size Nα.

$$MPC = \max_t P_c(t;\gamma,CLV,\varphi,\phi) \quad (9)$$

All the parameters in MPC are deterministic except the probability of true churners γ accepting the offer. Therefore, Verbraken et al. in [210] proposed the Expected Maximum Profit Measure (EMP), a probabilistic metric for customer churn.

- *Expected maximum profit (EMP)* - It is an extension of MPC, which assumes that the probability of true churners γ who accept the offer is a random variable that follows the beta distribution.

$$EMP = \int_\gamma P_c\big(T(\gamma);CLV,\varphi,c\big).h(\gamma)d\gamma \quad (10)$$

Here, $h(\gamma)$ denotes the probability density function for γ, and T(γ) is the optimal threshold that maximises the profit for a given value of γ.

Profit-based metrics such as MPC and EMP are used to determine the most effective predictive model for a business by considering profitability. They are designed to overcome the limitations of conventional performance metrics and are used to assess the performance of a classifier that maximises profitability and retention of profitable customers [31], [37], [79], [209]. Unlike the AUC, which measures the overall performance of a classifier, profit-based metrics evaluate a segment of customers that maximises the profit of a retention campaign. This approach is particularly useful for businesses looking to improve their retention campaigns by targeting their most profitable customers.

VI. GAP ANALYSIS AND RECOMMENDATIONS

This article provides a comprehensive guide for practitioners (such as marketing strategists, data analysts, and team members from the marketing analytics and business intelligence division), offering an in-depth review of machine learning approaches for customer churn prediction. The review synthesises understanding of the subject by surveying 212 articles, including 185 primary studies published from 2015 to January 2024, and drawing context from 27 supportive articles. This encompasses a detailed exploration of feature engineering, feature selection, model development,

TABLE 6. Summary of relevant literature reviews and their contributions to customer churn prediction.

Ref.	Targeted Industry & Dataset	Objective	Methodology	Features Type	Features Selection Strategy	Sampling	Remarks
[37]	Telecom (1 Public, 8 Private)	To optimise the profitability of DT	Integrate profit-based metrics into DT construction	N/A	N/A	N/A	Profit-driven DT are more effective in achieving business objectives compared to classic accuracy driven DT.
[36]	Telecom (Private)	To predict churn severity levels and generate retention campaigns	Fuzzy-based classifier and categorisation on usage patterns for retention (Voice, data, SMS)	CDR, complaints data	Domain knowledge & PCA	Stratified sampling	Fuzzy-based classifier performs better on noisy and incomplete data.
[59]	Online Games (Public)	Task1: To classify binary churn problem. Task 2: To predict survival weeks	Feature engineering was crucial for task 1, while tree-based and regression models were effective for task 2	In-game activity logs	Different teams used various strategies	N/A	Task 1: Teams used classifiers; LSTM performed best. Task 2: Teams used linear and tree-based approaches; tree-based approach performed best.
[40]	Telecom (Private)	To evaluate the performance of relational and non-relational classifiers	Empirical comparison of classifiers with significance tests	Network features, RFM, relational learner scores	N/A	N/A	Relational learner improves predictive power, suggesting wider churn influence beyond immediate neighbours.
[71]	Telecom (Private)	To combine behavioural and structural information considering different temporal granularities	Used tcc2vec, a window-based time-sliced approach on RFM-augmented call graph with representation learning	Behavioural interaction and structural call graph features	Feature extracted by representation learning	N/A	RFM-augmented networks with adapted node2vec are efficient and have better predictive performance than traditional RFM features.
[45]	Telecom (Public)	To predict churn, segment customers, and identify churning factors	EDA, Predictive modelling, factor analysis and customers segmentation	Different features per dataset	Chi-square test	SMOTE	Bayesian regression used for churn factor analysis.
[91]	Telecom (Public)	To improve the performance of churn prediction models	Bagging and Boosting ensembles	Various (demographic, CDR, data usage)	N/A	Under bagging	Boosting to strengthen base classifier; bagging for majority voting in the final classifier.
[65]	Telecom (Private)	To predict early churn using daily dynamic behaviour	Deep learning and feature-based approach on daily dynamic behaviour as a multivariate time series	Usage, KPI records, RFM	Automated feature engineering and statistic model	N/A	Daily behaviour-based models more accurate than monthly; statistical features not useful.
[66]	Telecom (Private)	To identify influential customers and extract dynamic behavioural patterns in call networks	Social network analysis with centrality metrics applied to time-ordered and aggregated graphs	Behavioural features (CDR)	N/A	N/A	Most of the static centrality metrics outperforms the temporal ones.
[46]	Financial (Private)	To study the relationship between spending behaviour, diversity of financial choices[17], and churning activities	Exploited spatio-temporal and choice features to predict customer churn	Spatio-temporal and choice features, demographic features	N/A	SVM-SMOTE	Found statistical correlations between spatio-temporal mobility, choice patterns, and churning activities.
[35]	Multiple (Private)	To enhance the predictive power of interpretable models: LR and DT	Hybrid classification model: logit leaf model combining LR and DT	N/A	Built-in feature importance selection	Undersampling	The hybrid model outperforms standalone classifiers: logistic model trees and the RF.
[122]	Telecom (Private)	To maximise profits of churn prediction model while ensuring good performance on imbalanced class distributions	Incorporate profit-based metrics into model construction	N/A	Fisher Score	Random undersampling	Models trained on profit-based metrics are more profitable than standard classification approaches evaluated with profit measures.

[17] shopping categories, fund transfer mode.

TABLE 6. *(Continued.)* **Summary of relevant literature reviews and their contributions to customer churn prediction.**

[113]	Telecom and Financial (Public)	To improve the accuracy of the classifier and avoid overfitting	Hybrid model	Various	N/A	N/A	Hybrid model prevents overfitting by incorporating misclassified instances and improving interpretability.
[5]	Banking (Private)	To enhance prediction accuracy by leveraging experts' opinion	Soft data (experts' opinion) and hard (models' prediction) data fusion	Financial, behavioural and demographic	N/A	N/A	Incorporating domain expertise leads to a precise analysis of customer churn rate.
[69]	Online Games (Private)	To evaluate different combinations of classifiers and labelling approaches for churn prediction	Performance comparison using AUC, ROC, and Accuracy	User activity data features	Feature aggregation and reduction using random ferns	N/A	DTs and ANNs excel. The simple labelling approach has the highest quality scores, while sliding windows enhance applicability.
[96]	Telecom (Private)	To increase churn detection rate while requiring low computational complexity	Feature selection using chaotic salp swarm optimisation with fuzzy rule-based classifier	N/A	Chaotic salp swarm optimisation for feature selection	N/A	The proposed technique selects fewer features and outperforms the other compared techniques.
[95]	Telecom (9 Private, 2 Public)	To compare class imbalance techniques for customer churn prediction	Empirical performance on profit based evaluation metric using two classifiers	NA	Fisher Score	Multiple sampling techniques	Results show the impact of evaluation metrics on model performance.
[92]	Telecom (Public)	To predict early churn and perform retention process	Identify churning customers using swish RNN with heuristics optimisation-based feature selection strategy	CDR	Brownian movement combined with butterfly optimisation	N/A	The method outperforms other deep learning-based approaches based on accuracy, specificity, and sensitivity.
[42]	Telecom (Public)	To enhance the prediction ability in customer churn	Hybrid model combining PCA with SVM model and optimised SVM's parameter through heuristic algorithm	CDR	PCA	Random oversampling	DT and SVM performed equally well on accuracy; however, the hybrid algorithm is the winner.
[31]	7 Telecom (Public), 4 others	To compare performance of homogeneous and heterogeneous ensembles with standalone classifiers	Various statistical tests	Various (CDR, RFM)	Fisher score	N/A	Heterogeneous classifiers have better predictive performance; gradient boosting classifiers also show promising results.
[108]	Internet service provider (Private)	To improve the predictability of classifier on churn dataset	Ensemble of K-means classifier and PSO	Demographic and RFM	PCA	N/A	Improved classification by combining PCA, K-Means, and PSO. Prediction based on CRISP-DM method.
[87]	Telecom (Public)	To optimise hyperparameters of deep canonically correlated autoencoder to improve prediction performance	Deer hunting optimisation for parameter tuning	CDR	N/A	N/A	Method outperforms other state-of-the-art models.
[110]	Financial (Public)	To train simple, effective 2-layered and 3-layered CNN churn prediction models with financial data heat maps	Data representation using fading channel patch-based heat map	Demographic, transactions, customer reviews, and purchases	N/A	N/A	Two-layered CNN model shows better performance.
[74]	Financial (Private)	To address data scarcity, high-class imbalance, and seasonality in churn predicting with limited historical data	Sliding multi-time slicing based training and testing framework	RFM	N/A	N/A	The method increases the training dataset by considering the same customers during multiple time slices.
[111]	Telecom (Public)	To improve customer retention through goal-oriented customer segmentation	Goal-oriented meta-heuristic modification of the ant colony DT	Signed-up services, financial variables, demographic	N/A	N/A	The classifier can be steered towards recall or precision based on the decision-maker's requirements.

TABLE 6. ***(Continued.)*** **Summary of relevant literature reviews and their contributions to customer churn prediction.**

[77]	Financial (Private)	To build reliable churn model with minimal business specific data and evaluate historical data's impact	Three established models (LR, R, and SVM) with a multi-slicing approach	Invoice data	Permutation-based feature importance approach	N/A	Models using longer spans of historical data tend to perform better.
[56]	Telecom (Public)	To improve the predictive performance of imbalanced customer churn dataset	Ensemble (Improve SMOTE and WELM with metaheuristic optimisation method)	CDR	N/A	Oversampling	Improved SMOTE sampling rate and WELM parameter tuning using rain optimisation algorithm characteristics.
[89]	Telecom (Public)	To select salient features from customer churn datasets	Feature selection approach combining two meta-heuristic algorithms	Various	Meta-heuristic optimisation (ACO-RSA)	N/A	ACO-RSA achieves a balanced exploration and exploitation rate and avoids getting trapped in local optima.
[125]	Financial and telecom (public)	To select the best features for distinguishing classes and eliminate unnecessary features	Feature reduction methods (Autoencoders, PCA, t-SNE, Xgboost, LDA)	Various	PCA, LDA, t-SNE, Xgboost	N/A	Xgboost algorithm performs the best among the feature reduction algorithms.
[112]	Telecom (Public)	To investigate the ability of ensembles to predict customer churn	Comparative analysis	Demographic, Subscribed services, payment method, length of contract	N/A	SMOTE	Combination of various classifiers and voting techniques outperforms individual classifiers.
[120]	Financial (Private)	To identify best classifier for financial (unique) dataset	Empirical performance evaluation	Transactions data	N/A	Undersampling	RF outperforms other classifiers.
[93]	Telecom (Public)	To design a predictive model that can adapt changes in customer behaviour	Adaptive learning (Naïve Bayes classifier with a Genetic Algorithm based feature weighting approach)	CDR	N/A	N/A	Feature weighting improves predictive performance.
[90]	Telecom (Public)	To analyse the performance of statistical classifiers and metaheuristic optimisation	Performance comparison using various evaluation metrics on imbalanced datasets	CDR	Metaheuristic optimisation	Random undersampling	Metaheuristics churn prediction has best performance.
[94]	Telecom (Public)	To compare the predictive performances of classifiers on various telecom datasets	Performance comparison based on traditional evaluation metrics	Demographic, CDR and financial	N/A	N/A	RF performed best on both balanced and imbalanced datasets.
[97]	Telecom (Public)	To improve robustness and effectiveness of churn prediction model	Ensemble combining the best performing classifiers	CDR, demographic and credit level	Analysis through Scatter plots, bar charts and cross-tabulations	N/A	Combines the best performing classifiers.
[121]	Financial (Public)	To improve the model performance on imbalanced dataset	Focal loss function was incorporated into Light-GBM to balance between classes	Demographic and transaction	N/A	N/A	Focal loss function improves the performance on metrics such as G-means and true positive rate.

and evaluation metrics, particularly emphasising the integration of profit-based measures for evaluating predictive performance. The key areas this review has focused on are summarised as follows.

The first focus was exploring the features relevant to churn prediction across diverse industries, with insights into the processes of their engineering and selection.

The second focus was understanding the trend behind machine learning methods and techniques in customer churn prediction.

The third focus was reviewing and synthesising which metrics are used for evaluating machine-learned models for churn prediction.

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

1) *Limited availability of up-to-date public datasets* - Most of the existing datasets are rather old or private [211]. 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.
2) *Lack of consensus for feature-set* - Some studies suggest using behavioural features. In contrast, other studies suggest using customer feedback and network features while developing a churn prediction model [11].
3) *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 the 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 generalisation across industry domains [106].
4) *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 [70], [203].
5) *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 [212].

Here are refined recommendation for addressing the identified gaps:

1) *Creation of novel public datasets* - To prioritise developing up-to-date, high-quality public datasets, companies should focus on data anonymisation, 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.
2) *Integration of feature-sets* - While modelling churn prediction, it is recommended i) To combine behavioural 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. Analysing feedback allows the identification of specific issues or areas of improvement.

TABLE 7. The list of the abbreviations used in this article.

Abbreviations	Description
ANN	Artificial Neural Networks
BaBaG	Boosted Bagging
BoBaG	Bagged Bagging
BNNGA	Bagging of Neural Network with learning based on Genetic Algorithm
CDR	Call Detail Record
CDRN	Class Distribution Relational Classifier
FRNN	Fuzzy-Rough Nearest Neighbours
FNN	Fuzzy Neural Networks
GBT	Gradient Boosted Trees
GBM	Gradient Boosting Machine
KNN	K-Nearest Neighbours
LSTM	Long Short Term Memory
MPM	Maximum Probability Machine
NLB	Network-only Link-based Classifier
OWANN	Ordered Weighted Average Nearest Neighbour
PCA	Principle Component Analysis
PSO	Particle Swarm Optimisation
RFM	Recency, Frequency, and Monetary value
ROC	Receiver Operating Characteristics
RNN	Recurrent Neural Network
RSA	Reptile Search Algorithm
RFID	Recency, frequency, importance and duration
SVM	Support Vector Machines
SHAP	SHapley Additive explanations
SPA-RC	Spreading Activation Relational Classifier
VQNN	Vaguely Quantified Nearest Neighbours
WVRN	Weighted Vote Relational Classifier

3) *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. Generally, the class of ensembles and deep learning perform better than simple methods, which is also evident from Figure 6.
4) *Design modelling techniques with profit-based metrics* - To incorporate profit-based metrics like Maximum Profit Criterion (MPC) and Expected Maximum Profit (EMP) into existing model techniques. This is crucial for developing churn prediction models tailored to retention efforts and maximisation of profitability.
5) *Adoption of explainable methods* - To use explainable methods from eXplainable AI (XAI) [188] for the interpretation of churn prediction models. This can support the computation and analysis of feature importance, discover bias within data, and help deliver transparent explanations to stakeholders, increasing trust.

APPENDIX

See Tables 5–7.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest associated with this paper.

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] T. Dierkes, M. Bichler, and R. Krishnan, "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks," *Decis. Support Syst.*, vol. 51, no. 3, pp. 361–371, Jun. 2011.

[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. 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

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

[22] P. S. H. Tan, Y. O. Choong, and I.-C. Chen, "The effect of service quality on behavioural 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.

[23] 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.

[24] N. Willys, "Customer satisfaction, switching costs and customer loyalty: An empirical study on the mobile telecommunication service," *Amer. J. Ind. Bus. Manage.*, vol. 8, no. 4, pp. 1022–1037, 2018.

[25] A. A. Q. Ahmed and D. Maheswari, "Churn prediction on huge telecom data using hybrid firefly based classification," *Egyptian Informat. J.*, vol. 18, no. 3, pp. 215–220, Nov. 2017.

[26] N. Vilkaite-Vaitone and I. Skackauskiene, "Service customer loyalty: An evaluation based on loyalty factors," *Sustainability*, vol. 12, no. 6, p. 2260, Mar. 2020.

[27] W. Reinartz, J. S. Thomas, and V. Kumar, "Balancing acquisition and retention resources to maximize customer profitability," *J. Marketing*, vol. 69, no. 1, pp. 63–79, Jan. 2005.

[28] N. Hariri, "Relevance ranking on Google: Are top ranked results really considered more relevant by the users?" *Online Inf. Rev.*, vol. 35, no. 4, pp. 598–610, Aug. 2011.

[29] K. Eria and B. P. Marikannan, "Systematic review of customer churn prediction in the telecom sector," *J. Appl. Technol. Innov.*, vol. 2, no. 1, pp. 1–14, 2018.

[30] L. Geiler, S. Affeldt, and M. Nadif, "A survey on machine learning methods for churn prediction," *Int. J. Data Sci. Analytics*, vol. 14, no. 3, pp. 217–242, Sep. 2022.

[31] M. Bogaert and L. Delaere, "Ensemble methods in customer churn prediction: A comparative analysis of the state-of-the-art," *Mathematics*, vol. 11, no. 5, p. 1137, Feb. 2023.

[32] P. Tabesh, E. Mousavidin, and S. Hasani, "Implementing big data strategies: A managerial perspective," *Bus. Horizons*, vol. 62, no. 3, pp. 347–358, May 2019.

[33] J. Moeyersoms and D. Martens, "Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector," *Decis. Support Syst.*, vol. 72, pp. 72–81, Apr. 2015.

[34] K. Eria and B. P. Marikannan, "Significance-based feature extraction for customer churn prediction data in the telecom sector," *J. Comput. Theor. Nanoscience*, vol. 16, no. 8, pp. 3428–3431, Aug. 2019.

[35] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *Eur. J. Oper. Res.*, vol. 269, no. 2, pp. 760–772, Sep. 2018.

[36] M. Azeem and M. Usman, "A fuzzy based churn prediction and retention model for prepaid customers in telecom industry," *Int. J. Comput. Intell. Syst.*, vol. 11, no. 1, p. 66, 2018.

[37] S. Höppner, E. Stripling, B. Baesens, S. V. Broucke, and T. Verdonck, "Profit driven decision trees for churn prediction," *Eur. J. Oper. Res.*, vol. 284, no. 3, pp. 920–933, Aug. 2020.

[38] K. W. De Bock and A. De Caigny, "Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling," *Decis. Support Syst.*, vol. 150, Nov. 2021, Art. no. 113523.

[39] S. K. Kwak and J. H. Kim, "Statistical data preparation: Management of missing values and outliers," *Korean J. Anesthesiology*, vol. 70, no. 4, p. 407, 2017.

[40] M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen, "Social network analytics for churn prediction in telco: Model building, evaluation and network architecture," *Expert Syst. Appl.*, vol. 85, pp. 204–220, Nov. 2017.

[41] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access*, vol. 4, pp. 7940–7957, 2016.

[42] B. Durkaya Kurtcan and T. Ozcan, "Predicting customer churn using grey wolf optimization-based support vector machine with principal component analysis," *J. Forecasting*, vol. 42, no. 6, pp. 1329–1340, Sep. 2023.

[43] J. Vijaya and E. Sivasankar, "Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector," *Computing*, vol. 100, no. 8, pp. 839–860, Aug. 2018.

[44] C. Rao, Y. Xu, X. Xiao, F. Hu, and M. Goh, "Imbalanced customer churn classification using a new multi-strategy collaborative processing method," *Expert Syst. Appl.*, vol. 247, Aug. 2024, Art. no. 123251.

[45] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated churn prediction and customer segmentation framework for Telco business," *IEEE Access*, vol. 9, pp. 62118–62136, 2021.

[46] E. Kaya, X. Dong, Y. Suhara, S. Balcisoy, and B. Bozkaya, "Behavioral attributes and financial churn prediction," *EPJ Data Sci.*, vol. 7, no. 1, p. 41, Dec. 2018.

[47] O. Soleiman-garmabaki and M. H. Rezvani, "Ensemble classification using balanced data to predict customer churn: A case study on the telecom industry," *Multimedia Tools Appl.*, vol. 83, no. 15, pp. 44799–44831, Oct. 2023.

[48] W. H. Khoh, Y. H. Pang, S. Y. Ooi, L.-Y.-K. Wang, and Q. W. Poh, "Predictive churn modeling for sustainable Bus. In the telecommunication industry: Optimized weighted ensemble machine learning," *Sustainability*, vol. 15, no. 11, p. 8631, May 2023.

[49] S. J. Haddadi, A. Farshidvard, F. D. S. Silva, J. C. dos Reis, and M. da Silva Reis, "Customer churn prediction in imbalanced datasets with resampling methods: A comparative study," *Expert Syst. Appl.*, vol. 246, Jul. 2024, Art. no. 123086.

[50] S. Saha, C. Saha, M. Mahidul Haque, M. Golam Rabiul Alam, and A. Talukder, "ChurnNet: Deep learning enhanced customer churn prediction in telecommunication industry," *IEEE Access*, vol. 12, pp. 4471–4484, 2024.

[51] P. Jiang, Z. Liu, L. Zhang, and J. Wang, "Hybrid model for profit-driven churn prediction based on cost minimization and return maximization," *Expert Syst. Appl.*, vol. 228, Oct. 2023, Art. no. 120354.

[52] B. Zhu, C. Qian, S. vanden Broucke, J. Xiao, and Y. Li, "A bagging-based selective ensemble model for churn prediction on imbalanced data," *Expert Syst. Appl.*, vol. 227, Oct. 2023, Art. no. 120223.

[53] A. Idris, A. Iftikhar, and Z. U. Rehman, "Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling," *Cluster Comput.*, vol. 22, no. S3, pp. 7241–7255, May 2019.

[54] P. Kate, V. Ravi, and A. Gangwar, "FinGAN: Chaotic generative adversarial network for analytical customer relationship management in banking and insurance," *Neural Comput. Appl.*, vol. 35, no. 8, pp. 6015–6028, Mar. 2023.

[55] J. B. G. Brito, G. B. Bucco, R. Heldt, J. L. Becker, C. S. Silveira, F. B. Luce, and M. J. Anzanello, "A framework to improve churn prediction performance in retail banking," *Financial Innov.*, vol. 10, no. 1, pp. 1–29, Jan. 2024.

[56] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, M. Elhoseny, and K. Shankar, "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," *Complex Intell. Syst.*, vol. 9, no. 4, pp. 3473–3485, Aug. 2023.

[57] R. Wirth and J. Hipp, "Crisp-DM: Towards a standard process model for data mining," in *Proc. 4th Int. Conf. practical Appl. Knowl. Discovery Data Mining*, vol. 1, 2000, pp. 29–39.

[58] D. Do, P. Huynh, P. Vo, and T. Vu, "Customer churn prediction in an Internet service provider," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2017, pp. 3928–3933.

[59] E. Lee, Y. Jang, D. Yoon, J. Jeon, S.-I. Yang, S.-K. Lee, D.-W. Kim, P. P. Chen, A. Guitart, P. Bertens, Á. Periáñez, F. Hadiji, M. Müller, Y. Joo, J. Lee, I. Hwang, and K.-J. Kim, "Game data mining competition on churn prediction and survival analysis using commercial game log data," *IEEE Trans. Games*, vol. 11, no. 3, pp. 215–226, Sep. 2019.

[60] W. Yang, T. Huang, J. Zeng, L. Chen, S. Mishra, and Y. E. Liu, "Utilizing players' playtime records for churn prediction: Mining playtime regularity," *IEEE Trans. Games*, vol. 14, no. 2, pp. 153–160, Jun. 2022.

[61] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," *Appl. Sci.*, vol. 11, no. 11, p. 4742, May 2021.

[62] M. R. Ismail, M. K. Awang, M. N. A. Rahman, and M. Makhtar, "A multi-layer perceptron approach for customer churn prediction," *Int. J. Multimedia Ubiquitous Eng.*, vol. 10, no. 7, pp. 213–222, Jul. 2015.

[63] S. Kim, D. Choi, E. Lee, and W. Rhee, "Churn prediction of mobile and online casual games using play log data," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180735.

[64] A. Perišić and M. Pahor, "Clustering mixed-type player behavior data for churn prediction in mobile games," *Central Eur. J. Oper. Res.*, vol. 31, no. 1, pp. 165–190, Mar. 2023.

[65] N. Alboukaey, A. Joukhadar, and N. Ghneim, "Dynamic behavior based churn prediction in mobile telecom," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113779.

[66] L. Calzada-Infante, M. Óskarsdóttir, and B. Baesens, "Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts," *Expert Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113553.

[67] M. R. Chayjan, T. Bagheri, A. Kianian, and N. G. Someh, "Using data mining for prediction of retail banking customer's churn behaviour," *Int. J. Electron. Banking*, vol. 2, no. 4, p. 303, 2020.

[68] E. Loria and A. Marconi, "Exploiting limited players' behavioral data to predict churn in gamification," *Electron. Commerce Res. Appl.*, vol. 47, May 2021, Art. no. 101057.

[69] K. Rothmeier, N. Pflanzl, J. A. Hüllmann, and M. Preuss, "Prediction of player churn and disengagement based on user activity data of a freemium online strategy game," *IEEE Trans. Games*, vol. 13, no. 1, pp. 78–88, Mar. 2021.

[70] E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang, and H. K. Kim, "Profit optimizing churn prediction for long-term loyal customers in online games," *IEEE Trans. Games*, vol. 12, no. 1, pp. 41–53, Mar. 2020.

[71] S. Mitrović, B. Baesens, W. Lemahieu, and J. De Weerdt, "Tcc2vec: RFM-informed representation learning on call graphs for churn prediction," *Inf. Sci.*, vol. 557, pp. 270–285, May 2021.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] S. Maldonado, G. Domínguez, D. Olaya, and W. Verbeke, "Profit-driven churn prediction for the mutual fund industry: A multisegment approach," *Omega*, vol. 100, Apr. 2021, Art. no. 102380.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] I. V. Pustokhina, D. A. Pustokhin, A. Rh, T. Jayasankar, C. Jeyalakshmi, V. G. Díaz, and K. Shankar, ''Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms,'' *Inf. Process. Manage.*, vol. 58, no. 6, Nov. 2021, Art. no. 102706.
[89] I. Al-Shourbaji, N. Helian, Y. Sun, S. Alshathri, and M. A. Elaziz, ''Boosting ant colony optimization with reptile search algorithm for churn prediction,'' *Mathematics*, vol. 10, no. 7, p. 1031, Mar. 2022.
[90] J. Vijaya and 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, Sep. 2019.
[91] S. Tavassoli and H. Koosha, ''Hybrid ensemble learning approaches to customer churn prediction,'' *Kybernetes*, vol. 51, no. 3, pp. 1062–1088, Feb. 2022.
[92] R. Sudharsan and E. N. Ganesh, ''A swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy,'' *Connection Sci.*, vol. 34, no. 1, pp. 1855–1876, Dec. 2022.
[93] A. Amin, A. Adnan, and S. Anwar, ''An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes,'' *Appl. Soft Comput.*, vol. 137, Apr. 2023, Art. no. 110103.
[94] D. D. Adhikary and D. Gupta, ''Applying over 100 classifiers for churn prediction in telecom companies,'' *Multimedia Tools Appl.*, vol. 80, nos. 28–29, pp. 35123–35144, Nov. 2021.
[95] B. Zhu, B. Baesens, and S. K. L. M. vanden Broucke, ''An empirical comparison of techniques for the class imbalance problem in churn prediction,'' *Inf. Sci.*, vol. 408, pp. 84–99, Oct. 2017.
[96] J. Faritha Banu, S. Neelakandan, B. T. Geetha, V. Selvalakshmi, A. Umadevi, and E. O. Martinson, ''Artificial intelligence based customer churn prediction model for Bus. Markets,'' *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–14, Sep. 2022.
[97] Y. Liu, J. Fan, J. Zhang, X. Yin, and Z. Song, ''Research on telecom customer churn prediction based on ensemble learning,'' *J. Intell. Inf. Syst.*, vol. 60, no. 3, pp. 759–775, Jun. 2023.
[98] M. Fathian, Y. Hoseinpoor, and B. Minaei-Bidgoli, ''Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods,'' *Kybernetes*, vol. 45, no. 5, pp. 732–743, May 2016.
[99] R. Yu, X. An, B. Jin, J. Shi, O. A. Move, and Y. Liu, ''Particle classification optimization-based BP network for telecommunication customer churn prediction,'' *Neural Comput. Appl.*, vol. 29, no. 3, pp. 707–720, Feb. 2018.
[100] A. Amin, B. Shah, A. M. Khattak, F. J. Lopes Moreira, G. Ali, A. Rocha, and S. Anwar, ''Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods,'' *Int. J. Inf. Manage.*, vol. 46, pp. 304–319, Jun. 2019.
[101] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, ''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.
[102] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq, and Y. S. Lee, ''Transfer learning and meta classification based deep churn prediction system for telecom industry,'' 2019, *arXiv:1901.06091*.
[103] M. M. Uner, F. Guven, and S. T. Cavusgil, ''Churn and loyalty behavior of Turkish digital natives: Empirical insights and managerial implications,'' *Telecommun. Policy*, vol. 44, no. 4, May 2020, Art. no. 101901.
[104] R. Soltani, M. Ashrafi, M. M. S. Esfahani, and H. Farvaresh, ''Competitive pricing of complementary telecommunication services with subscriber churn in a duopoly,'' *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121447.
[105] M. A. S. Thorat and V. R. Sonawane, ''A random forest churn prediction model: An study of machine learning techniques for churn prediction and factor identification in the telecom sector,'' *Scandin. J. Inf. Syst.*, vol. 35, no. 1, pp. 818–824, 2023.
[106] N. Gordini and V. Veglio, ''Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry,'' *Ind. Marketing Manage.*, vol. 62, pp. 100–107, Apr. 2017.
[107] K. Hasumoto and M. Goto, ''Predicting customer churn for platform businesses: Using latent variables of variational autoencoder as consumers' purchasing behavior,'' *Neural Comput. Appl.*, vol. 34, no. 21, pp. 18525–18541, Nov. 2022.
[108] M. Sadeghi, M. N. Dehkordi, B. Barekatain, and N. Khani, ''Improve customer churn prediction through the proposed PCA-PSO-K means algorithm in the communication industry,'' *J. Supercomput.*, vol. 79, no. 6, pp. 6871–6888, Apr. 2023.
[109] P. P. Šimović, C. Y. T. Chen, and E. W. Sun, ''Classifying the variety of customers' online engagement for churn prediction with a mixed-penalty logistic regression,'' *Comput. Econ.*, vol. 61, no. 1, pp. 451–485, Jan. 2023.
[110] Seema and G. Gupta, ''Development of fading channel patch based convolutional neural network models for customer churn prediction,'' *Int. J. Syst. Assurance Eng. Manage.*, vol. 15, no. 1, pp. 391–411, Jan. 2024.
[111] J. Kozak, K. Kania, P. Juszczuk, and M. Mitręga, ''Swarm intelligence goal-oriented approach to data-driven innovation in customer churn management,'' *Int. J. Inf. Manage.*, vol. 60, Oct. 2021, Art. no. 102357.
[112] Y. Beeharry and R. Tsokizep Fokone, ''Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry,'' *Concurrency Computation, Pract. Exper.*, vol. 34, no. 4, p. e6627, Feb. 2022.
[113] A. M A and S. K K, ''An efficient hybrid classifier model for customer churn prediction,'' *Int. J. Electron. Telecommun.*, vol. 69, pp. 11–18, Dec. 2022.
[114] Z. Saberi, O. K. Hussain, and M. Saberi, ''Data-driven personalized assortment optimization by considering customers' value and their risk of churning: Case of online grocery shopping,'' *Comput. Ind. Eng.*, vol. 182, Aug. 2023, Art. no. 109328.
[115] Y. Suh, ''Machine learning based customer churn prediction in home appliance rental business,'' *J. Big Data*, vol. 10, no. 1, p. 41, Apr. 2023.
[116] M. R. Flores-Méndez, M. Postigo-Boix, J. L. Melús-Moreno, and B. Stiller, ''A model for the mobile market based on customers profile to analyze the churning process,'' *Wireless Netw.*, vol. 24, no. 2, pp. 409–422, Feb. 2018.
[117] S. Saleh and S. Saha, ''Customer retention and churn prediction in the telecommunication industry: A case study on a Danish university,'' *Social Netw. Appl. Sci.*, vol. 5, no. 7, p. 173, Jul. 2023.
[118] M. Milošević, N. Živić, and I. Andjelković, ''Early churn prediction with personalized targeting in mobile social games,'' *Expert Syst. Appl.*, vol. 83, pp. 326–332, Oct. 2017.
[119] Q.-F. Wang, M. Xu, and A. Hussain, ''Large-scale ensemble model for customer churn prediction in search ads,'' *Cognit. Comput.*, vol. 11, no. 2, pp. 262–270, Apr. 2019.
[120] R. A. de Lima Lemos, T. C. Silva, and B. M. Tabak, ''Propension to customer churn in a financial institution: A machine learning approach,'' *Neural Comput. Appl.*, vol. 34, no. 14, pp. 11751–11768, Jul. 2022.
[121] J. Li, X. Bai, Q. Xu, and D. Yang, ''Identification of customer churn considering difficult case mining,'' *Systems*, vol. 11, no. 7, p. 325, Jun. 2023.
[122] S. Maldonado, J. López, and C. Vairetti, ''Profit-based churn prediction based on minimax probability machines,'' *Eur. J. Oper. Res.*, vol. 284, no. 1, pp. 273–284, Jul. 2020.
[123] S. P. Kumar and N. Leema, ''PCP: Profit-driven churn prediction using machine learning techniques in banking sector,'' *Int. J. Performability Eng.*, vol. 19, no. 5, p. 303, 2023.
[124] A. Khodadadi, S. A. Hosseini, E. Pajouheshgar, F. Mansouri, and H. R. Rabiee, ''Choracle: A unified statistical framework for churn prediction,'' *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 4, pp. 1656–1666, Apr. 2022.
[125] S. M. Sina Mirabdolbaghi and B. Amiri, ''Model optimization analysis of customer churn prediction using machine learning algorithms with focus on feature reductions,'' *Discrete Dyn. Nature Soc.*, vol. 2022, pp. 1–20, Jun. 2022.
[126] C. K. Leung, A. G. M. Pazdor, and J. Souza, ''Explainable artificial intelligence for data science on customer churn,'' in *Proc. IEEE 8th Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2021, pp. 1–10.
[127] A. K. Ahmad, A. Jafar, and K. Aljoumaa, ''Customer churn prediction in telecom using machine learning in big data platform,'' *J. Big Data*, vol. 6, no. 1, pp. 1–24, Dec. 2019.
[128] G. Marín Díaz, R. A. Carrasco, and D. Gómez, ''RFID: A fuzzy linguistic model to manage customers from the perspective of their interactions with the contact center,'' *Mathematics*, vol. 9, no. 19, p. 2362, Sep. 2021.
[129] G. Marín Díaz, J. J. Galán, and R. A. Carrasco, ''XAI for churn prediction in B2B models: A use case in an enterprise software company,'' *Mathematics*, vol. 10, no. 20, p. 3896, Oct. 2022.

[130] N. N. Y. Vo, S. Liu, X. Li, and G. Xu, "Leveraging unstructured call log data for customer churn prediction," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106586.

[131] A. De Caigny, K. Coussement, K. W. De Bock, and S. Lessmann, "Incorporating textual information in customer churn prediction models based on a convolutional neural network," *Int. J. Forecasting*, vol. 36, no. 4, pp. 1563–1578, Oct. 2020.

[132] K. A. Tarnowska and Z. Ras, "NLP-based customer loyalty improvement recommender system (CLIRS2)," *Big Data Cognit. Comput.*, vol. 5, no. 1, p. 4, Jan. 2021.

[133] F. v. Wangenheim, N. V. Wünderlich, and J. H. Schumann, "Renew or cancel? Drivers of customer renewal decisions for IT-based service contracts," *J. Bus. Res.*, vol. 79, pp. 181–188, Oct. 2017.

[134] C. Vairetti, I. Aránguiz, S. Maldonado, J. P. Karmy, and A. Leal, "Analytics-driven complaint prioritisation via deep learning and multicriteria decision-making," *Eur. J. Oper. Res.*, vol. 312, no. 3, pp. 1108–1118, Feb. 2024.

[135] S. De and P. Prabu, "A representation-based query strategy to derive qualitative features for improved churn prediction," *IEEE Access*, vol. 11, pp. 1213–1223, 2023.

[136] J. Wang, S. Shi, X. Wang, and J. Yu, "Image analysis considering textual correlations enables accurate user switching tendency prediction," *Optoelectronics Lett.*, vol. 19, no. 8, pp. 498–505, Aug. 2023.

[137] J. Vijaya and E. Sivasankar, "A study of feature selection techniques for predicting customer retention in telecommunication sector," *Int. J. Bus. Inf. Syst.*, vol. 31, no. 1, p. 1, 2019.

[138] E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens, and M. Snoeck, "Profit maximizing logistic model for customer churn prediction using genetic algorithms," *Swarm Evol. Comput.*, vol. 40, pp. 116–130, Jun. 2018.

[139] A. Perisic and M. Pahor, "Extended RFM logit model for churn prediction in the mobile gaming market," *Croatian Oper. Res. Rev.*, vol. 11, no. 2, pp. 249–261, 2020.

[140] B. Garimella, G. V. S. N. R. V. Prasad, and M. H. M. K. Prasad, "Churn prediction using optimized deep learning classifier on huge telecom data," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 3, pp. 2007–2028, Mar. 2023.

[141] X. Xiahou and Y. Harada, "B2C e-commerce customer churn prediction based on K-means and SVM," *J. Theor. Appl. Electron. Commerce Res.*, vol. 17, no. 2, pp. 458–475, Apr. 2022.

[142] P. Ramesh, J. Jeba Emilyn, and V. Vijayakumar, "Hybrid artificial neural networks using customer churn prediction," *Wireless Pers. Commun.*, vol. 124, no. 2, pp. 1695–1709, May 2022.

[143] D. AL-Najjar, N. Al-Rousan, and H. AL-Najjar, "Machine learning to develop credit card customer churn prediction," *J. Theor. Appl. Electron. Commerce Res.*, vol. 17, no. 4, pp. 1529–1542, Nov. 2022.

[144] S. O. Abdulsalam, M. O. Arowolo, Y. K. Saheed, and J. O. Afolayan, "Customer churn prediction in telecommunication industry using classification and regression trees and artificial neural network algorithms," *Indonesian J. Electr. Eng. Informat. (IJEEI)*, vol. 10, no. 2, pp. 431–440, Jun. 2022.

[145] D. Hason Rudd, H. Huo, and G. Xu, "Improved churn causal analysis through restrained high-dimensional feature space effects in financial institutions," *Hum.-Centric Intell. Syst.*, vol. 2, nos. 3–4, pp. 70–80, Jul. 2022.

[146] T. Zhang, S. Moro, and R. F. Ramos, "A data-driven approach to improve customer churn prediction based on telecom customer segmentation," *Future Internet*, vol. 14, no. 3, p. 94, Mar. 2022.

[147] M. Z. Abedin, P. Hajek, T. Sharif, M. S. Satu, and M. I. Khan, "Modelling bank customer behaviour using feature engineering and classification techniques," *Res. Int. Bus. Finance*, vol. 65, Apr. 2023, Art. no. 101913.

[148] P. Jiang, Z. Liu, M. Z. Abedin, J. Wang, W. Yang, and Q. Dong, "Profit-driven weighted classifier with interpretable ability for customer churn prediction," *Omega*, vol. 125, Jun. 2024, Art. no. 103034.

[149] F. Wu, F. Lyu, J. Ren, P. Yang, K. Qian, S. Gao, and Y. Zhang, "Characterizing Internet card user portraits for efficient churn prediction model design," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1–17, Feb. 2024.

[150] Z. Wu, L. Jing, B. Wu, and L. Jin, "A PCA-AdaBoost model for e-commerce customer churn prediction," *Ann. Oper. Res.*, vol. 4, pp. 1–18, Jan. 2022.

[151] L. Zhao, Q. Gao, X. Dong, A. Dong, and X. Dong, "K-local maximum margin feature extraction algorithm for churn prediction in telecom," *Cluster Comput.*, vol. 20, no. 2, pp. 1401–1409, Jun. 2017.

[152] S. A. Panimalar and A. Krishnakumar, "Customer churn prediction model in cloud environment using DFE-WUNB: ANN deep feature extraction with weight updated tuned naive Bayes classification with block-Jacobi SVD dimensionality reduction," *Eng. Appl. Artif. Intell.*, vol. 126, Sep. 2023, Art. no. 107015.

[153] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection," *PLoS ONE*, vol. 17, no. 12, Dec. 2022, Art. no. e0278095.

[154] K. Coussement, S. Lessmann, and G. Verstraeten, "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry," *Decis. Support Syst.*, vol. 95, pp. 27–36, Mar. 2017.

[155] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: A machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, Feb. 2022.

[156] Y. Chen, L. Zhang, Y. Zhao, and B. Xu, "Implementation of penalized survival models in churn prediction of vehicle insurance," *J. Bus. Res.*, vol. 153, pp. 162–171, Dec. 2022.

[157] J. Nagaraju and J. Vijaya, "Boost customer churn prediction in the insurance industry using meta-heuristic models," *Int. J. Inf. Technol.*, vol. 14, no. 5, pp. 2619–2631, Aug. 2022.

[158] N. Jajam, N. P. Challa, K. S. L. Prasanna, and C. H. V. S. Deepthi, "Arithmetic optimization with ensemble deep learning SBLSTM-RNN-IGSA model for customer churn prediction," *IEEE Access*, vol. 11, pp. 93111–93128, 2023.

[159] R. Chinnaraj, "Bio-inspired approach to extend customer churn prediction for the telecom industry in efficient way," *Wireless Pers. Commun.*, vol. 133, no. 1, pp. 15–29, Nov. 2023.

[160] I. AlShourbaji, N. Helian, Y. Sun, A. G. Hussien, L. Abualigah, and B. Elnaim, "An efficient churn prediction model using gradient boosting machine and metaheuristic optimization," *Sci. Rep.*, vol. 13, no. 1, p. 14441, Sep. 2023.

[161] E. Sivasankar and 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, Nov. 2019.

[162] A. Wangperawong, C. Brun, O. Laudy, and R. Pavasuthipaisit, "Churn analysis using deep convolutional neural networks and autoencoders," 2016, *arXiv:1604.05377*.

[163] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, Feb. 2019, Art. no. 022022.

[164] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. D. Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, R. Jiang, H. Khosravi, F. Lecue, G. Malgieri, A. Páez, W. Samek, J. Schneider, T. Speith, and S. Stumpf, "Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions," 2023, *arXiv:2310.19775*.

[165] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Commun. ACM*, vol. 63, no. 1, pp. 68–77, Dec. 2019.

[166] M. Kiguchi, W. Saeed, and I. Medi, "Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest," *Appl. Soft Comput.*, vol. 118, Mar. 2022, Art. no. 108491.

[167] T. T. Dang, K. N. Hoang, L. B. Thanh, T. N. T. Thuy, and C. N. Quoc, "Constructing and understanding customer spending prediction models," *Social Netw. Comput. Sci.*, vol. 4, no. 6, p. 852, Nov. 2023.

[168] E.-B. Lee, J. Kim, and S.-G. Lee, "Predicting customer churn in mobile industry using data mining technology," *Ind. Manage. Data Syst.*, vol. 117, no. 1, pp. 90–109, Feb. 2017.

[169] M. Pejić Bach, J. Pivar, and B. Jaković, "Churn management in telecommunications: Hybrid approach using cluster analysis and decision trees," *J. Risk Financial Manage.*, vol. 14, no. 11, p. 544, Nov. 2021.

[170] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, Jun. 2015.

[171] E. M. Elgohary, M. Galal, A. Mosa, and G. A. Elshabrawy, "Smart evaluation for deep learning model: Churn prediction as a product case study," *Bull. Electr. Eng. Informat.*, vol. 12, no. 2, pp. 1219–1225, Apr. 2023.

[172] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, ''Customer churn prediction in telecommunication industry using data certainty,'' *Int. J. Scientific Res. Sci., Eng. Technol.*, vol. 94, pp. 252–258, Feb. 2020.
[173] M. T. Quasim, A. Sulaiman, A. Shaikh, and M. Younus, ''Blockchain in churn prediction based telecommunication system on climatic weather application,'' *Sustain. Comput., Informat. Syst.*, vol. 35, Sep. 2022, Art. no. 100705.
[174] M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie, ''Comprehensive review on twin support vector machines,'' *Ann. Oper. Res.*, vol. 2, pp. 1–46, Mar. 2022.
[175] S. J, C. Gangadhar, R. K. Arora, P. N. Renjith, J. Bamini, and Y. D. Chincholkar, ''E-commerce customer churn prevention using machine learning-based business intelligence strategy,'' *Meas., Sensors*, vol. 27, Jun. 2023, Art. no. 100728.
[176] W. Etaiwi, M. Biltawi, and G. Naymat, ''Evaluation of classification algorithms for banking customer's behavior under apache spark data processing system,'' *Proc. Comput. Sci.*, vol. 113, pp. 559–564, 2017.
[177] R. M. Wahul, A. P. Kale, and P. N. Kota, ''An ensemble learning approach to enhance customer churn prediction in telecom industry,'' *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 9s, pp. 258–266, 2023.
[178] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, ''Customer churn prediction in telecom sector using machine learning techniques,'' *Results Control Optim.*, vol. 14, Mar. 2024, Art. no. 100342.
[179] P. P. Singh, F. I. Anik, R. Senapati, A. Sinha, N. Sakib, and E. Hossain, ''Investigating customer churn in banking: A machine learning approach and visualization app for data science and management,'' *Data Sci. Manage.*, vol. 7, no. 1, pp. 7–16, Mar. 2024.
[180] Y. Zhou, W. Chen, X. Sun, and D. Yang, ''Early warning of telecom enterprise customer churn based on ensemble learning,'' *PLoS ONE*, vol. 18, no. 10, Oct. 2023, Art. no. e0292466.
[181] R. P. Sari, F. Febriyanto, and A. C. Adi, ''Analysis implementation of the ensemble algorithm in predicting customer churn in Telco data: A comparative study,'' *Informatica*, vol. 47, no. 7, pp. 1–18, Jul. 2023.
[182] K. Arik, M. Gezer, and S. T. Tayali, ''The study of indicators affecting customer churn in MMORPG games with machine learning models,'' *Upravlenets*, vol. 13, no. 6, pp. 70–85, Jan. 2023.
[183] M. Óskarsdóttir, K. E. Gísladóttir, R. Stefánsson, D. Aleman, and C. Sarraute, ''Social networks for enhanced player churn prediction in mobile free-to-play games,'' *Appl. Netw. Sci.*, vol. 7, no. 1, p. 82, Dec. 2022.
[184] A. Y. W. Chong, K. W. Khaw, W. C. Yeong, and W. X. Chuah, ''Customer churn prediction of telecom company using machine learning algorithms,'' *J. Soft Comput. Data Mining*, vol. 4, no. 2, pp. 1–22, Oct. 2023.
[185] B. Prabadevi, R. Shalini, and B. R. Kavitha, ''Customer churning analysis using machine learning algorithms,'' *Int. J. Intell. Netw.*, vol. 4, pp. 145–154, Aug. 2023.
[186] P. Swetha and R. Dayananda, ''A customer churn prediction model in telecom industry using improved_XGBoost,'' *Int. J. Cloud Comput.*, vol. 12, nos. 2–4, pp. 277–294, 2023.
[187] G. Vilone, L. Rizzo, and L. Longo, ''A comparative analysis of rule-based, model-agnostic methods for explainable artificial intelligence,'' in *Proc. 28th Irish Conf. Artif. Intell. Cogn. Sci.*, vol. 2771, 2020, pp. 85–96.
[188] G. Vilone and L. Longo, ''A quantitative evaluation of global, rule-based explanations of post-hoc, model agnostic methods,'' *Frontiers Artif. Intell.*, vol. 4, p. 160, Nov. 2021.
[189] J. Tao, Y. Xiong, S. Zhao, R. Wu, X. Shen, T. Lyu, C. Fan, Z. Hu, S. Zhao, and G. Pan, ''Explainable AI for cheating detection and churn prediction in online games,'' *IEEE Trans. Games*, vol. 15, no. 2, pp. 242–251, May 2022, doi: 10.1109/TG.2022.3173399.
[190] E. Domingos, B. Ojeme, and O. Daramola, ''Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector,'' *Computation*, vol. 9, no. 3, p. 34, Mar. 2021.
[191] J. Xu, J. Liu, T. Yao, and Y. Li, ''Prediction and big data impact analysis of telecom churn by backpropagation neural network algorithm from the perspective of Bus. Model,'' *Big Data*, vol. 11, no. 5, pp. 355–368, Oct. 2023.
[192] O. Çelik and U. O. Osmanoglu, ''Comparing to techniques used in customer churn analysis,'' *J. Multidisciplinary Develop.*, vol. 4, no. 1, pp. 30–38, 2019.
[193] L. Saha, H. K. Tripathy, T. Gaber, H. El-Gohary, and E.-S.-M. El-kenawy, ''Deep churn prediction method for telecommunication industry,'' *Sustainability*, vol. 15, no. 5, p. 4543, Mar. 2023.
[194] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, ''Customer churn prediction using composite deep learning technique,'' *Sci. Rep.*, vol. 13, no. 1, p. 17294, Oct. 2023.
[195] G. Z. Khan, I. Ulhaq, I. Adil, S. Ulhaq, and I. Ullah, ''Privacy-preserving based technique for customer churn prediction in telecom industry,'' *VFAST Trans. Softw. Eng.*, vol. 11, no. 3, pp. 73–80, Sep. 2023.
[196] A. Rodan, A. Fayyoumi, H. Faris, J. Alsakran, and O. Al-Kadi, ''Negative correlation learning for customer churn prediction: A comparison study,'' *Sci. World J.*, vol. 2015, pp. 1–7, Oct. 2015.
[197] Y. Liu, M. Shengdong, G. Jijian, and N. Nedjah, ''Intelligent prediction of customer churn with a fused attentional deep learning model,'' *Mathematics*, vol. 10, no. 24, p. 4733, Dec. 2022.
[198] P. Routh, A. Roy, and J. Meyer, ''Estimating customer churn under competing risks,'' *J. Oper. Res. Soc.*, vol. 72, no. 5, pp. 1138–1155, May 2021.
[199] I. Ullah, A. Rios, V. Gala, and S. Mckeever, ''Explaining deep learning models for tabular data using layer-wise relevance propagation,'' *Appl. Sci.*, vol. 12, no. 1, p. 136, Dec. 2021.
[200] A. Perišić, D. Š. Jung, and M. Pahor, ''Churn in the mobile gaming field: Establishing churn definitions and measuring classification similarities,'' *Expert Syst. Appl.*, vol. 191, Apr. 2022, Art. no. 116277.
[201] M. Chaudhary, L. Gaur, N. Jhanjhi, M. Masud, and S. Aljahdali, ''Envisaging employee churn using MCDM and machine learning,'' *Intell. Autom. Soft Comput.*, vol. 33, no. 2, pp. 1009–1024, 2022.
[202] F. Devriendt, D. Moldovan, and W. Verbeke, ''A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics,'' *Big Data*, vol. 6, no. 1, pp. 13–41, Mar. 2018.
[203] T. Verbraken, W. Verbeke, and B. Baesens, ''Profit optimizing customer churn prediction with Bayesian network classifiers,'' *Intell. Data Anal.*, vol. 18, no. 1, pp. 3–24, Jan. 2014.
[204] V. Umayaparvathi and K. Iyakutti, ''A survey on customer churn prediction in telecom industry: Datasets, methods and metrics,'' *Int. Res. J. Eng. Technol. (IRJET)*, vol. 3, no. 4, pp. 1–49, 2016.
[205] M. Hassouna, A. Tarhini, T. Elyas, and M. S. Abou Trab, ''Customer churn in mobile markets: A comparison of techniques,'' *Int. Bus. Res.*, vol. 8, no. 6, pp. 1–12, May 2015.
[206] F. Movahedi, R. Padman, and J. F. Antaki, ''Limitations of receiver operating characteristic curve on imbalanced data: Assist device mortality risk scores,'' *J. Thoracic Cardiovascular Surgery*, vol. 165, no. 4, pp. 1433–1442, Apr. 2023.
[207] A. Lemmens and S. Gupta, ''Managing churn to maximize profits,'' *Marketing Sci.*, vol. 39, no. 5, pp. 956–973, Sep. 2020.
[208] R. Jafari-Marandi, J. Denton, A. Idris, B. K. Smith, and A. Keramati, ''Optimum profit-driven churn decision making: Innovative artificial neural networks in telecom industry,'' *Neural Comput. Appl.*, vol. 32, no. 18, pp. 14929–14962, Sep. 2020.
[209] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, ''New insights into churn prediction in the telecommunication sector: A profit driven data mining approach,'' *Eur. J. Oper. Res.*, vol. 218, no. 1, pp. 211–229, Apr. 2012.
[210] T. Verbraken, W. Verbeke, and B. Baesens, ''A novel profit maximizing metric for measuring classification performance of customer churn prediction models,'' *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 961–973, May 2013.
[211] A. Bugajev, R. Kriauzienė, O. Vasilecas, and V. Chadyšas, ''The impact of churn labelling rules on churn prediction in telecommunications,'' *Informatica*, vol. 33, no. 2, pp. 247–277, 2022.
[212] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. D. Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, R. Jiang, H. Khosravi, F. Lecue, G. Malgieri, A. Páez, W. Samek, J. Schneider, T. Speith, and S. Stumpf, ''Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions,'' *Inf. Fusion*, vol. 106, Jun. 2024, Art. no. 102301.

AWAIS MANZOOR is currently pursuing the Ph.D. degree in business technology with Technological University Dublin (TU Dublin), Ireland. Previously, he was a Lecturer with the Department of Computer Science, University of Lahore, Pakistan. He has authored several journal articles and book chapters. His research interests include business analytics, machine learning, image processing, and deep learning, particularly in the explainability and interpretability of AI-based decision-making systems. He is a Reviewer of journals, such as IEEE Systems Journal.

M. ATIF QURESHI received the B.S. degree in computer science from the University of Karachi, Pakistan, the M.S. degree in computer science from Korea Advanced Institute of Science and Technology, South Korea, and the joint Ph.D. degree in computer science from the National University of Ireland Galway, Ireland, and the University of Milano-Bicocca, Italy. Currently, he is a Lecturer/Assistant Professor and the Research Leader of the eXplainable Analytics Research Group (XAG), Technological University Dublin, Ireland. His research interests include natural language processing, machine learning, disinformation space, and explainable artificial intelligence. As a Principal Investigator and the Technical Lead, he has significantly contributed to diverse projects funded by Science Foundation Ireland, the Irish Research Council, Enterprise Ireland, and the EU. Notably, he has successfully licensed an outcome to a prominent media organization in Ireland, showcasing his dedication to impactful applied research in interdisciplinary domains.

ETAIN KIDNEY received the M.A. degree in higher education, the M.Sc. degree in entrepreneurship, and the Ph.D. degree in inclusive entrepreneurship. She is currently the Head of the School of Marketing and Entrepreneurship, TU Dublin, where she has been a Researcher and a Lecturer for 18 years. Her research interests include sustainability, business technology, and inclusive practice.

LUCA LONGO received the bachelor's and master's degrees in computer science, statistics, and health informatics and the Ph.D. degree in artificial intelligence from Trinity College Dublin and the master's degree in pedagogy from Technological University Dublin. He is currently the Leader of the Artificial Intelligence and Cognitive Load Research Laboratories. With his team of doctoral and postdoctoral scholars, he conducts fundamental research in eXplainable Artificial Intelligence, defeasible reasoning, and non-monotonic argumentation. He also performs applied research in machine learning and predictive data analytics, mainly applied to the problem of mental workload modeling using electroencephalography. He is also the Founder of the World Conference on eXplainable Artificial Intelligence. He is actively engaged in the dissemination of scientific material to the public, contributing to the non-profit TED Organization.