



A Decade of Churn Prediction Techniques in the TelCo Domain: A Survey

Annalisa Barsotti¹ · Gabriele Gianini² · Corrado Mio³ · Jianyi Lin⁴ · Himanshi Babbar⁵ · Aman Singh⁶ · Fatma Taher⁷ · Ernesto Damiani¹

Received: 28 December 2023 / Accepted: 14 February 2024
© The Author(s) 2024

Abstract

This work surveys the research contributions of the last decade to the prediction of customer churn and adds a perspective toward what is yet to be reached. The main objective of this article is to report on (1) the methods and algorithms studied, the evaluation metrics adopted, and the results achieved, (2) the data used, and (3) the issues and limitations identified. Furthermore, the work highlights the gaps in the current literature and suggests a direction for future research.

Keywords Telecom customer churn · Churn prediction · Machine learning techniques · Causal inference

Introduction

The term *Customer Churn* refers to the departure of customers from a service company. This is also known as customer attrition, defection, or turnover. This phenomenon, which is physiological for any service business, is of great relevance to companies, and its quantification enters into one of the typical key performance indicators, the churn rate. Customer churn has a substantial impact on many aspects of a company's business and entails opportunity costs: not

only do company's sales decrease, but it is also not easy to compensate by attracting new customers. In fact, acquiring a new customer is usually 5–6 times, or even up to 20 times [1], more expensive than retaining an existing one [2]. For this reason, companies adopt customer retention strategies, and most of them rely on the ability to predict which customers are at risk of departure. It is common sense that a customer can change service provider due to the availability of a wide range of services/products in a competitive market [3]. This is why churn is an even harder challenge in the TelCo domain: this market has evolved into an increasingly

This article is part of the topical collection “Digital Ecosystems” guest edited by Richard Chbeir and Yannis Manolopoulos.

✉ Annalisa Barsotti
annalisa.barsotti@unimi.it

Gabriele Gianini
gabriele.gianini@unimib.it

Corrado Mio
corrado.mio@ku.ac.ae

Jianyi Lin
jianyi.lin@unicatt.it

Himanshi Babbar
himanshi.babbar@chitkara.edu.in

Aman Singh
amansingh.x@gmail.com

Fatma Taher
fatma.taher@zu.ac.ae

Ernesto Damiani
ernesto.damiani@unimi.it

- ¹ Dipartimento di Informatica (DI), Università degli Studi di Milano, via Celoria 18, 20133 Milan, Italy
- ² Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo), Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milan, Italy
- ³ Emirates Innovation Center (EBTIC), Khalifa University of Science and Technology, Hadbat Al Zaafran, 127788 Abu Dhabi, UAE
- ⁴ Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Largo A. Gemelli 1, 20123 Milan, Italy
- ⁵ Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India
- ⁶ Universidad Internacional Iberoamericana, Arecibo, PR 00613, USA
- ⁷ College of Technological Innovation, Zayed University, Academic City, 19282 Dubai, UAE

competitive ecosystem in the last few decades, and the diffusion of smartphones has made this evolution more erratic.

Well-established expert knowledge tells that the churn rate increases when the monthly consumption decreases and the customer satisfaction is low. However, factors such as high dependency on products, high package value, and bundle contracts decrease departure. Price is a key factor in customer turnover based on the assumption of the same quality of services (provided by other companies). Building a corporate brand is also important. Different special products can be launched for various market segments to improve the dependence of customers on products. Another strategy relies on the identification of "high value" customers: their value, their index of contribution, their needs and complaints help the company put more resources, improve satisfaction, and adopt the convergence business strategy that increases the dependency on the products and the transfer cost for abandoning the network. In addition, efficient complaints handling plays a key role in winning back customers. However, the problem of the customer churn prediction (CCP) cannot be addressed only by using well-established domain expert knowledge: there is the opportunity to exploit the potential wealth represented by customer-related data. The abundance, fine-grained character, and diversity of those data, with the consequent difficulty for humans to detect characteristic patterns, calls for the use of methods that go beyond simple data analysis and bring the challenge into the arena of machine learning. Machine learning is a branch of artificial intelligence that uses inductive methods to learn models from the data: a predictive churn model could be learned, in principle, by looking at a number of customers traces ending in a defection. Depending on the available data and the precise definition of the problem, many techniques could be applied for this purpose; moreover, considering the additional business requirements, many different challenges could arise, in addition to the obvious requirement of prediction accuracy, the requirement of trustworthiness and interpretability of the learned model, to mention a few. We did not consider comparing the results in a systematic way as it is not possible to make a comparison if the data sets are too heterogeneous.

Contributions

This work surveys the literature on machine learning methods for CCP within the scope of the TelCo domain in the approximate period of the last decade, which has seen the establishment and the diffusion of smartphones. Due to continuous changes in the dynamics of the market, the present work complements other surveys, published in 2013 [4], 2017 [5], and 2022 [6]. We will briefly comment on them in the following.

The main contributions of this survey are the following: (1) we review the machine learning techniques applied to the churn prediction problem, focusing on the techniques and on the specification of data they use. (2) The machine learning models are presented and analyzed, describing the capabilities of the model, the list of assumptions involved in each model, and other relevant aspects. (3) We outline future research directions for the field and discuss their implications.

Article Structure

The paper is organized as follows. After introducing the customer churn prediction problem in this section, in Sect. 2 we present the methodology used to select the articles and the pattern we adopted to analyze each article. In Sect. 3, we survey the articles first discussing the existing surveys and then the non-survey papers. In Sect. 4, we discuss the identified problems and open challenges and point to future research directions.

Acronyms

In order to ensure a consistent presentation of concepts and models, a list of acronyms used in the text is provided in Table 1.

Literature Review Methodology

Research Scope and Questions

The survey discusses the topic of prediction and notes that research papers often attempt to address related questions when discussing this aspect. The survey pertains to the issue of prediction. If such information is available, research papers may also attempt to tackle other associated inquiries. We searched for works that addressed the following: What are the most effective machine learning algorithms in customer churn prediction (according to a variety of metrics)?; considering a predictive algorithm, given a customer at risk, What factors make him/her more likely to churn?; Is it possible to quantify the effect of possible interventions on those variables that are actionable from the point of view of the company? A precision: a particular set of techniques in customer fidelization and retention is that of recommender systems; they aim to tailor the service offer according to the specific customer profile; this area of research falls outside the scope of this work.

Table 1 The acronyms that are used in the survey

AUC	Area under ROC curve
ANN	Artificial neural network
BN	Bayesian network
CART	Classification and regression tree
CCP	Customer churn prediction
DL	Deep learning
DT	Decision tree
GBM	Gradient boosted machine
GP	Genetic programming
KNN	K-nearest neighbors
LDA	Linear discriminant analysis
LR	Linear regression
LoR	Logistic regression
MCC	Matthews correlation coefficient
NB	Naïve Bayes
NN	Neural network
NPS	Net promoter score
PCA	Principal component analysis
PCALB	Principal component analysis load balancing
POL	Polynomial kernel
PSM	Propensity score matching
PSO	Particle swarm optimization
RA	Regression analysis
RBF	Radial basis function
RF	Random forest
SIG	Sigmoid kernel
SRI	Single rule induction
SMO	Sequential minimal optimization
SNA	Social network analysis
SOM	Self-organizing maps
SVM	Support vector machine
TelCo	Telecommunication

Article Selection Process

We searched “Google Scholar” as it offers a number of options to combine multiple search terms with boolean operators. The basic search string used was: “customer” AND “churn” AND “prediction” AND (“telco” OR “telecom”); we restricted the search to the last 15 years.

Inclusion Criteria

The papers selected in our survey (in addition to the English language criterion) meet the following criteria: journal papers published on first quartile or the second quartile journals according to the Scopus ranking; the classes’ conferences are A++, A+, A and B according to the core ranking; studies that analyze any customer churn prediction aspect related to the TelCo domain; studies with at least two

citations; literature review papers, summarizing results of the previous studies; papers that describe machine learning techniques applied to customer churn prediction.

Exclusion Criteria

We did not consider book chapters, work-in-progress articles, posters, theses; short versions of works already considered; discussion papers.

The final number of papers considered was 27.

Article Report Schema

The elements that we report about each paper include algorithms, datasets, methods, metrics, and results:

- *Algorithms and methods* the main focus of the survey of each paper. We show the relation between techniques and articles in the summary Table 2.
- *Datasets* real data relating to TelCo customers are often incomplete, noisy, and unstructured. For each paper, we report synthetically on the data used.
- *Metrics* several performance metrics were proposed in the literature. Some works do not consider only the usual metrics, such as precision and recall, but also cost-related metrics.
- *Results* the results presented in each research article are discussed and compared.

Articles Description

Most works frame the problem of prediction of churn in the form of a classification task with two classes (churn/non-churn). The use of traditional classification methods such as decision trees, random forest, classification and regression tree, logistic regression, naïve Bayesian classifiers, K-nearest neighbor clustering, support vector machines, is very common, but artificial neural networks, self-organizing maps and evolutionary algorithms are also widely used. A common issue to be faced is the class imbalance problem: for obvious reasons, the churn class is underrepresented in the typical data. Several articles analyze the different impacts of type I errors (false positives) and type II errors (false negatives) on the business when predicting churn.

Survey Articles

The survey by Hashmi et al. [7] focuses on churn in fixed-line services. It highlights the role of customer call details, demographic, complaints, billing information, and contractual data for the mobile domain. The most used classification techniques are DT, ANN, LoR, in conjunction

Table 2 Comparison table of algorithms including the surveys

Algorithm	Papers
ANN	[1, 4, 6, 7, 9, 14, 22, 25, 26, 29, 31, 35, 43–46]
AdaBoost	[17, 33]
BN	[11, 14, 17, 28]
CART	[41]
CatBoost	[33]
Cluster analysis	[7]
DTs	[1, 6, 7, 9, 11, 13, 22, 25, 28, 29, 31, 33, 40, 45, 47, 48]
Extra tree	[33]
GBoost machine tree	[43, 47]
Genetic programming	[17]
K-nearest neighbor	[6, 26, 28, 29, 31, 41, 43, 44, 46]
K-means clustering	[25, 45]
LDA	[41]
Literature review	[4–7]
LR	[31, 44, 45]
LoR	[1, 1, 6, 7, 13, 24, 29, 33–35, 40–42, 45, 49, 50]
Logit boost	[44]
NB	[1, 6, 11, 17, 26, 29, 31, 33, 35, 43, 44]
PCALB	[44]
PSM	[40]
PSO	[28]
RF	[6, 17, 22, 33, 35, 40, 47, 49, 50]
Regression analysis	[1]
RotBoost	[17]
Rotation forest	[17, 31]
Self organizing maps	[8]
SRI	[43]
Support vector machines	[1, 6, 8, 12, 14, 22, 26, 31, 33, 35, 44, 49, 50]
XGBoost	[6, 33, 44, 47, 50]

with cluster analysis. Among the other challenges faced by the analyst are data confidentiality, missing data, incomplete or very large data sets containing noise can affect the reliability of the predictive model.

The survey by Garcia et al. [5], considers the application of knowledge extraction in the form of business analytics. In particular, the use of intelligent data analysis, or data mining, for the analysis of market-surveyed information can be of great assistance for churn management. An in-depth discussion of churn is also proposed in the context of customer continuity management. The study by Kamalraj et al. [4] focuses on interpreting existing models from a management perspective, so as to guide the investment in appropriate efforts considering the customer's lifetime value. It points to the considerable number of false positives and false negatives and raises the issue of an excessive retention effort both toward customers who are not leaving but are erroneously considered churners and towards customers who are not considered churners, but in fact are real churners.

The work of Geiler et al. [6] is both a survey and a comparative study covering churn in different business domains. The authors use publicly available churn data sets from several fields (TelCo, music streaming, human resources, newspapers). In addition, by framing the problem of churn as an anomaly detection problem, they also try the relevant techniques on a Bank Marketing data set, a Credit Card Fraud Detection data set and a thyroid disease data set, which they categorize as *churn-like*. They consider semi-supervised techniques (iForest and DevNet) and supervised algorithms (DT, RF, KNN, NB, LoR, SVM with RBF kernel (SVM-RBF) and without kernel (SVM), XGBoost (an optimized distributed gradient boosting library), ANN and generalized extreme value neural network (GEV-ANN)) in association with different undersampling, oversampling, and hybrid sampling strategies, to compensate for the class imbalance when the issue is not already addressed by the algorithm (GEV-ANN, iForest, and DevNet are specifically designed for imbalance binary classification or anomaly detection). They eventually focus on an ensemble method

based on the four main best-performing approaches: LR, XGBoost, RF, and ANN. Based on the algorithms' performance results, they recommend using the ensemble LR, XGBoost, and RF with no sampling to analyze novel churn-like data sets. The work focuses on predicting churn and does not investigate the causes of defection or discusses actionable interventions.

Research Articles

Tsai et al. [8], use a hybrid technique combining ANN techniques and SOMs. They considered a Customer Relationship Management (CRM) data set provided by American TelCo companies, containing 51,306 subscribers, including 34,761 churners and 16,545 non-churners, from July 2001 to January 2002. Customers which had been with the company for at least six months are defined as mature customers. Churn is calculated on the basis of whether the subscriber left the company during the period 31–60 days after the subscriber was originally sampled. The authors combined two ANNs (ANN+ANN) and ANN+SOM. The first part of the hybrid model filters out unrepresentative data or outliers and evaluates a prediction model. The hybrid models outperform the single ANN in terms of accuracy and type I and type II errors. The ANN+ANN model performs better and is more stable.

In the research by Umayaparvathi et al. [9] machine learning techniques (DT and ANN) are investigated. The data set is acquired from PAKDD 2006 data mining competition and aggregated for 6 months. A total of 24,000 customers is used and 252 attributes (almost 50% are derived attributes). Data are aggregated over six months: customer behavior during the previous 6 months is used to predict churners during the next month (7th month). DT is observed to outperform the ANN model. For attribute selection, they use information gain and entropy. Customer demography, bill and payment, call detail record, customer service, turn out to be the most important types of features for the prediction. As evaluation metrics the authors use accuracy and error rate: for the DT, the accuracy is 98.88%, the error rate is 1.11%, the false positive rate is 0.93% and the false negative rate is 2.23%. For the ANN the precision is 98.43%, the false positive rate is 1.26%, and the false negative rate is 3.40% and the error rate is 1.56%.

The article by Qureshi et al. [10] studies the following machine learning techniques: LR, LoR, ANN, KNN, DT, CART, Chi-squared automatic interaction detector (CHAID), exhaustive CHAID and quick, unbiased, and efficient statistical tree (QUEST).

The data set used comes from a TelCo operator with approximately 106,500 customers [100,264 active users (94.1%) and 6231 churners (5.9%)]. A customer is considered "Active" if he/she continues to use the network,

"Churner" in case the contract with the network is terminated. The data report about traffic type [outgoing, incoming, voice, SMS (short message service), traffic destination (on-net, competition), rate plan, loyalty, traffic behavior, etc. The authors address the problem of class imbalance by randomly oversampling the minority class by keeping the churners in an active ratio to 40:60 approximately].

The following evaluation metrics are considered: precision, recall, and F1-score. The authors also study variable importance by calculating the p value with respect to the target variable and the Spearman correlation to identify variables that are closely correlated with the customer's status (the main variables turn out to be: credit score; No. of penalties for non-payment; No. of outgoing calls to rival networks; No. of incoming SMS from rival networks; No. of days of outgoing activity). The results show that the exhaustive CHAID algorithm is the most accurate algorithm (accuracy about 70%). Recall for active users is 76.9%, and for churners it is 8.5% (an improvement of 68.5% over the best previous results. The recall in the test set for active customers is 76.3% and for churners it is 60.5%.

The study by Kirui et al. [11] focuses on feature selection to improve the classification. Data were obtained from a European TelCo company and were collected in a period of 3 months in 1997. The data set consists of 112 attributes and 106,405 instances of which 5.6% churners and remaining active subscribers. The authors consider phone traffic figures and customer profile data and extract a set of synthesis features that form the contract-related description, the call pattern description, and the call pattern changes description. The results of NB and BN are compared to those of C4.5 DT. The results show an improvement of prediction performance thanks to the new features for all models: higher true positive rates for the probabilistic classifiers and better overall accuracy for DT.

Brandusoiu et al. [12], propose a predictive model consisting of SVM with four kernel functions: a RBF kernel, a linear kernel (LIN), a polynomial kernel (POL) and sigmoid kernel (SIG). They focus on the importance of understanding and preparing the data set and then building and evaluating the models. The data set used is from the University of California, Irvine, Department of Information and Computer Science: it is a complete data set, and no attributes are missing. It contains call detail records and has 21 attributes for each of its 3333 subscribers. For each subscriber, they have information about their corresponding inbound / outbound call count, inbound / outbound SMS count, and voice mail.

The minority class is oversampled. The POL polynomial kernel function performs better (overall accuracy of 88.56%). RBF, LIN, and POL also have good performance (around 80%).

Vafeiadis et al. [1] present a comparative study on the most popular machine learning methods: ANN, SVM, DTs,

NB, regression analysis (RA), LoR. They first evaluate these classifiers through the usual cross-validation and then test their performance with boosting for the algorithms where this is possible (NB and LoR cannot be boosted because of the lack of free parameters to be tuned): to this purpose, they use the AdaBoost.M1 algorithm. Finally, Monte Carlo simulations are used to explore the models' parameter space and to determine the most effective approach. The data set originally from the UCI Machine Learning Repository, which is included in package C50 of the R language, contains 5000 samples.

Two of the best performing methods in terms of the corresponding testing error are the two-layer back-propagation ANN with 15 hidden units and the DT classifier; both methods achieve the accuracy of 94% and the F1-score of 77%. SVM classifiers (RBF and POL kernels) obtain an accuracy of about 93% and an approximate F1-score of 73%. The NB and LoR methods fail shortly with approximation accuracy 86% and an F1-score of approximately 53% and 14%. Using the AdaBoost.M1 algorithm determines an improvement in performance for the three remaining classifiers: the accuracy improves between 1% and 4%, while the F1 scores are between 4.5% and 15%. The best classifier is the boosted SVM (SVM-POL with AdaBoost) with an accuracy of almost 97% and an F1-score greater than 84%.

The study by Dahiya et al. [13] presents a new framework for the prediction of churn. A discussion of various prediction models, a comparison of quality measures such as regression analysis, and DTs are proposed. They use hybrid classification techniques to highlight a possible association between customer lifetime value and churn prediction. The data set for this study is acquired from the KDD Cup 2009. It is used to analyze the marketing tendency of customers from the large databases of the French TelCo company Orange with a total of 18,000 attributes. DT and LoR are used. Extensive experimentation shows that the accuracy achieved by the DT J48 is much higher than the one of LoR.

The article by Brandusoiu et al. [14] presents a new methodology to predict the probable churn of subscribers in the prepaid mobile sector by applying some ANN architectures to the call details record data set. The overall performance to predict between churning and non-churning is 99.55%. The prediction of churners on prepaid mobile devices is approximately 99%. Three kinds of machine learning algorithms are explored: ANN, SVM, (the parameters are learned by a divide-and-conquer approach: sequential minimal optimization (SMO) algorithm [15]; and by a fast SVM algorithm [16]) and BN [the structure is learned by iterative parent-child-based learning of Markov blanket (IPC-MB)]. The data set (provided by the University of California) consists of prepaid mobile call details records with 21 attributes each. It contains information on the usage of a mobile TelCo system and has

a total number of 3333 subscribers with 15 continuous and 5 discrete variables each, and the churn dependent variable with two classes Yes/No. For each subscriber, it contains information on the call count inbound/outbound, the SMS count inbound/outbound, and voice mail (call detail record). The recall and specificity of SVM are 100%, the recall for ANN and BN is approximately 99%, and the specificity is 100% for ANN and 99% for BN. The models have an overall precision of 99.10% for BN, 99.55% for ANN, and 99.70% for SVM. In general, the models perform well (from 99 to 100%).

The study by Idris et al. [17] proposes an approach based on the use of genetic programming (GP) search capabilities by an AdaBoost-based iterative approach (AB). It uses Orange and Cell2Cell data sets. The Orange data set comprises 50,000 instances where only 3276 are churners. Cell2Cell includes 40,000 instances where 20,000 are churners. The majority class of the Orange data set is undersampled using PSO to address the class imbalance. This method, combined with the GP-AdaBoost algorithm, provides the Ch-GPAB prediction system (where Ch stands for Churn). The recall and specificity measures or AUC curves are used. All the classifiers used (RF; rotation forest; RotBoost and GP-AdaBoost) show non-optimal performance on the Orange data set (imbalanced). However, the algorithms perform well on the Cell2Cell (balanced) data set, and GP-AdaBoost achieves higher prediction performance. Ch-GPAB achieves recall scores of 89% (Orange) and 93% (Cell2Cell), the best prediction performance (in terms of AUC) reported in the data sets. AUC obtained by GP-AdaBoost ranges from 70 to 91% over 30 independent simulations.

The AUC is studied in the Orange data set by making a comparison based on Ch-GPAB (0.751), gradient boost machine (0.737), decision stump-based model (0.725), DT-based model (0.715), a BN-based approach (0.714). They compare the performance of the Ch-GPAB (AUC: 0.910) and the NB-based approach (AUC: 0.818) on the Cell2Cell data set. A McNemar statistical test is performed to evaluate the confidence level of the Ch-GPAB system's prediction performance. A comparison of Ch-GPAB with Chr-PmRF (Churn particle swarm random forest) [18] and RUS-Boost (random under sampling boost) is made. McNemar's confusion matrices show that 10,200 instances are correctly predicted by the proposed Ch-GPAB, but incorrectly classified by Chr-PmRF for the Cell2Cell data set. A total of 1300 instances are correctly classified by Ch-GPAB and incorrectly predicted by Chr-PmRF for the Orange data set. Furthermore, 450 (Cell2Cell) and 22 (Orange) are incorrectly predicted by the proposed Ch-GPAB but correctly predicted by Chr-PmRF. The proposed Ch-GPAB correctly classified 3,125 instances, while only 11 instances are correctly classified by RUS-Boost and are incorrectly classified by Ch-GPAB in the Orange data set.

The article by Ahmad et al. [19] develops a model of churn prediction for customers who are likely to drop their subscription to Syriatel and MTN. They use DT, RF, gradient boosted machine and XGBoost algorithms. The data set includes all information from prepaid customers collected over 9 consecutive months. Customers considered churners are defined as being in the idle phase for at least 2 months. The remaining customers are labeled as active customers, except those acquired in the last 4 months. The sample includes 5 million customers: 0.3 million churned customers and 4.7 million active customers. Such a wide data set is valuable but also very challenging, due to a number of factors, among them: data volume; data variety (structured, semistructured, XML/JSON, unstructured; class unbalancing (the churn customers are about just 5%); number of features (10,000 columns (before preprocessing), including service, product, offer related to calls, SMS, MMS, Internet, personnel, and demographic information), missing values (some because not all customers have the same subscription). To cope with the data volume, the authors use a suitable big data platform, a data warehouse system. The data imbalance problem is faced by under-sampling or over-sampling without rebalancing. The authors leverage also social network related information based on the call details record of the last 4 months working out typical descriptors such as **degree centrality measures, similarity values, and customer network connectivity**. Feature engineering is based on both statistics features (i.e., aggregating for each customer the columns values per month for the numerical features and by counting distinct values for the categorical ones), and social network features are considered to build a **Social Network Graph**. The social network graph consists of the GSM (i.e. global system for mobile communication) number of subscribers (i.e. nodes) and the interactions (i.e., calls, SMS, and MMS) between subscribers (i.e. edges which are **directed**). The total social graph includes 15 million nodes that represent the phone numbers (i.e. SyriaTel, MTN, and baseline) and more than 2.5 Billion edges. The social graph is a weighted directed graph (i.e. the number of shared events between every two customers represents the edges' weight). Three graphs are generated depending on the used edges' weight (i.e. the normalized calling duration between customers; the normalized total number of calls, SMS, and MMS and the mean of the previous two normalized weights). To extract graph-based features, the weighted PageRank [20] and weighted Sender Rank [21] algorithms are used to produce two features for each graph that contain *Sink nodes* (i.e. nodes with zero outgoing and many incoming interactions; nodes with zero incoming and many outgoing interactions). In addition, other characteristics are considered. In particular: the degree of centrality, the number of distinct friends in receive and send behavior (i.e. IN and OUT degrees), neighbor connectivity (i.e. the

average connectivity of neighbors for each customer), local clustering coefficient (i.e. number of existing connections in neighborhood divided number of all possible connections: how close the customer's friends are). The Jaccard similarity and Cosine similarity are considered for each customer and help in finding similar customers in the network. The social network features are used for the first time to improve the prediction of churn, contributing positively to the accuracy of the prediction with the other statistical features. After both feature transformation and selection, the number of prediction features used is about 2,000. The results are the following: by using the **statistical** characteristics, using the last 6 months of the raw data set, the AUC is 84%; by using **SNA** characteristics, extracted from the last 4 months of that data set, the authors get AUC: 75.3%; by combining both SNA features and the statistical features, the results improve significantly, and the AUC reaches 93.3%. The article by Andrews et al. [22] studies the effectiveness of the RF, SVM and KNN algorithms in comparison with some DL models.

The data set comes from a TelCo Company in Belgium and contains a total of 22 factors from 10,000 customers (i.e. user's system history, day time call design, customer FICO rating, night call design, initial information plan, and current information plan): 29% of the customers are churners. They show that DL models perform similarly to conventional classifiers such as SVM and RF. The work by Apurva Sree et al. [23] considers algorithms such as SVM, RF, and LoR. The IBM Watson data set is used, covering 7000 customers, with 26.6% of customers who moved from one service provider to another. The accuracy obtained by the different algorithms is: LoR: 80.75%, RF: 80.88%, SVM: 82%. The authors find that the churn rate is higher for customers who have a month-to-month contract and that older citizens churn more than younger people. In the study by Shah et al. [24] a set of generic features is proposed that can be used for most non-subscription business settings for developing churn prediction systems. Furthermore, they apply **causal reasoning**.

The data are about sale and payment of a dealer, per day, per month, or on some other time granularity. A **dealer** is someone who acts as a broker between the business and the end user. The aim is to predict the churn of these dealers. The data contain sales and payment transactions from April 2016 to December 2018. Dealers who have not transacted since September 30th are considered churned. Sales data from 6000 customers are analyzed.

Generic features can be extracted from sales and payment data of almost all non-subscription-based businesses. The feature set is based on recency, frequency, and monetary analysis [a marketing technique used to quantitatively determine which customers are the most valuable ones by examining how recently a customer has purchased (recency), how often they purchase (frequency), and how much the customer spends (monetary)]. An ANN is trained for dealers who have

not done sales transactions in the last 120 days. The resulting test accuracy is 79.65%. A system for **causal analysis** of churn tries to identify a set of causes that may have led to churn of customers (and thus helps to derive customer retention strategies more effectively than simple prediction). The number of complaints; salesman or point of contact changed; orders canceled due to understock; returns due to defective material are mentioned to be the main **causal factors**. The authors propose two models specifically for causality analysis of customer churn: a **counterfactuals-based causal model** and a **cause prediction** approach based on the Bayes theorem.

The study by Sniegula et al. [25] is driven by research questions related to methodological class comparison, such as: “Is it really necessary to turn toward the complex ANNs?”; “Can satisfactory results be achieved with the use of simpler statistical approaches?”; “Is it enough to rely on simple unsupervised techniques, such as clustering, rather than supervised learning for classification?”. To answer these questions, they compare three machine learning techniques, that are K-means, DT, and ANN, by analyzing the data set “Churn in the Telecom Industry” taken from BigML, a managed platform offering Machine Learning as a service to share and develop data sets and models. The data set has 3333 records with 20 attributes (16 numerical, 1 textual, and 3 boolean attributes, one of which contains information about customer churn).

The K-means algorithm has been applied adopting four different distance measures: Euclidean, Manhattan, Chebyshev and Cosine distance, leading to a 60% accuracy and 30% F1-score as best results of the average performance measures. On the other hand, the ANN architectures with 3–5 linear layers, of 256 neurons each, were more favorably compared showing improved results up to 87.11% accuracy and 45.38% F1-score on 33 tests.

The best performances have been achieved by the CART model on 28 tests with an accuracy of 94.98% and an F1-score of 80.80%. However, as expected, the prediction was more accurate for non-churning customers, as evidenced by 80% precision in most tests, due to the class imbalance. Although they concluded that DT performed best, especially on data sets with mostly numerical features, the work suffers from issues concerning low significance of the experiments, such as few tests carried out.

The work by Amin et al. [26] presents a just-in-time (JIT) approach to provide useful answers to customer churn questions, especially when historical data are absent, such as at the initial stage of a new company or for those companies lacking or not having access to data archives. The method exploits the transfer learning mechanisms in that it uses publicly available data from two different companies: a large data set of 3333 records with 21 numerical features and another one of 5784 records with 250 input

variables from the KDD Cup 1999 competition. Indeed, from a methodological viewpoint, they aim at comparing individual classifiers or homogeneous ensemble classifiers based on bagging in JIT-CCP against heterogeneous ensembles based on stacking KNN, NB, ANN, and SVM base classifiers. They show the differences in feature space for the cross-company training set, considering the imbalanced data and the effect of rebalancing through random undersampling (RUS) method. The classifiers are hence trained on enough historical data stored in a company’s CRM system, and the fitted model is then applied to the data of a newly established company. Their experiments demonstrate that the best performance of ensemble-free methods, resulting in an average \pm standard deviation (in %) of 55.3 ± 7.13 in accuracy and 47.26 ± 11.12 in F1-score, can be actually improved with a homogeneous ensemble by 3.94% in accuracy and 9.47% in F1-score. Finally, the results reveals that their heterogeneous ensemble method, namely the JIT-CCP model, is the most effective, since it outperforms the simpler ones in accuracy by 18.03% and in F1-score by 14.69% while also reducing the misclassification rate to 18.03%. This strengthening effect is even obtained simply with the SVM base classifier in **heterogeneous ensemble**, comparing to homogeneous ensemble or individual classifier approach.

The study by Jolfoo et al. [27] briefly reviews machine learning techniques to reduce the churn rate in telecom companies and then proposes a hybrid approach based on KNN, which does not require prior knowledge of the distribution of data and ANN. In this pipeline, simple Pearson correlation coefficient estimates are applied to select relevant predictors among customer demographic variables, account, and line utilization variables. Despite the authors’ use of standard procedures to train the ANN with backpropagation and the KNN with properly tuned class boundaries in Euclidean space, their work lacks the full specification of the hybridization process and evident experimental results. However, they emphasize that company efforts must be devoted to clearly identifying the reasons behind customer behaviors.

Asghar et al. [28] face the churn prediction problem with a wrapper-based best feature selection approach using the PSO as search criterion in conjunction with a specific classifier among DT, NB, KNN and LoR. The experiments are carried out on a churn prediction data set of year 2020 with a sample size of 4250 and 19 characteristics, assessing the results by 10-fold cross-validation.

First, they show that the PSO-DT combination achieves the highest accuracy (94.56%) by selecting the best subset of features, namely 8 ones: international plan, total day minutes, total eve charge, number of vmail messages, total international charge, total night minutes, total international calls and number of customer service calls. The authors also evaluated the performances of all the methods in terms of accuracy against dimensionality reduction

(DR) achieved. A graphical comparison demonstrates that PSO-DT and PSO-NN combinations reduce the feature dimensions by more than 50% while still attaining a high accuracy of 94.56% and 89.2%, respectively. PSO-LoR selects a moderate number of features (DR = 42.1%), achieving a lower accuracy of 87.18%. Instead, PSO-NB performs the most conservative DR by only 26.32%, although it outperforms PSO-LoR in accuracy. Finally, the authors also suggested that their churn forecasting model could be applied to other industries characterized by highly competitive markets.

Bauer et al. [29] propose a novel method for customer lifetime value (CLV) prediction using a tailored DL approach. Despite the work is not focused on churn prediction, the authors claim that such a scenario in the telecommunication industry is a straightforward application of their methods. The proposed DL model consists of an encoder-decoder sequence-to-sequence recurrent neural network (RNN) with augmented temporal convolutions, combined then with gradient boosting machines and a set of novel features in a hybrid framework. The approach is evaluated on a private data set of more than 1 mln of customers' sales, named Children, provided by a European e-commerce company selling products for children and families, whose corresponding purchase data span a period of more than three years. Then, the findings are validated against a public, but smaller data set, called UKRetail, consisting of 541,909 purchase records from more than 4000 customers of a UK-based company over a period of about one year.

Other real-world data sets used in alternative CLV prediction methods include the CDNOW data set [30] from an online shopping application scenario that involves the sale of music. Notice that the time series data sets used in this CLV evaluation context usually lack information about churn, which is why the latter is hence modeled implicitly in the approach when the future CLV is predicted to be zero at some point in time.

Moreover, to overcome the data sparsity problem due to the abundance of attributes, mostly related to orders and customer behavior, the model learns an embedding of the recorded purchase logs to reliably perform clustering of similar customers.

Only high-importance features of the Children data set are selected to be used in the RNN model. The empirically determined models are applied to capture long-term trends in purchasing behavior, such as a steady increase in order volume over time and periodic patterns. The proposed method(s) outperform the four baseline alternatives that are auto-regressive moving average (ARMA), negative binomial distribution (NBD), Markov chain (MC) and embedding-based RF models, in terms of root mean square error and mean absolute error loss measures. In this comparison, the best values are obtained on both data sets with the stacked

combination of the gradient boosted machine (GBM) and the sequence-to-sequence RNN model (GBM-S2S).

The article by Gu et al. [31] compares the accuracy and efficiency of several commonly used algorithms: DT C4.5, CART, KNN, LR, SVM, NB (Gaussian NB, Bernoulli NB), RF, ANN. The reported data set includes 221,770 users, of which 11,282 (5%) are churn customers, with 37 features, including consumption data attributes, and consumption characteristic data, such as customer age, consumption, online time, online time period, online traffic, etc. The researchers found that the CART is fast, accurate (82%), and well suited to predict customer churn. The CART algorithm is simple and does not require a high level of technology, so it is suitable for data support personnel at all levels to quickly analyze big data for customer churn. They also apply a filtering by interquartile range as a quick method of clearing abnormal values, for simply and easily adjusting abnormal data (outliers) to the non-abnormal range. The original data set contains null values and extremely large values (Internet traffic and deposits) which need to be cleaned up. To test each algorithm separately, the data before and after data cleaning are compared. The split between train-test is 70–30% [221,770 pieces that were divided into 161,770 pieces of training data (around 73%) and 60,000 pieces of test data (around 27%)]. Features analysis and multi-feature analysis based on experience are performed and 37 features are selected. They analyze the accuracy of the train, the accuracy of the test, the AUC, and the run time of each algorithm. Using the DT, C4.5 algorithm, the performances achieve a test accuracy of 83% and an AUC of 78%.

Jain et al. [32] consider a number of machine learning algorithms for the prediction of churn in the TelCo, banking, and IT sectors: a hybrid model of DT and LoR; PCA with LoR and Logit Boost; CNN-VAE (convolutional neural network with variational autoencoder); LoR; logit boost; XGBoost and RF.

The study uses a publicly available Orange data set, which contains 3333 subscriber entries (483 churners) and 21 attributes. The best algorithm in the TelCo sector is XGBoost with a precision of 82.942%, in the banking sector RF with a precision of 86.312%, and in the IT sector LoR with a precision of 90.136%. The research by Lalwani et al. [33] studies the algorithms LoR, NB, SVM, DT, RF, XGBoost, CatBoost, AdaBoost classifier, and extra tree classifier. The results show that AdaBoost and XGBoost outperform the other algorithms in terms of performance measures [accuracy (81.71% and 80.8%, respectively), precision, recall, F1-score, AUC score (84%, is achieved by both)].

Research by Zhao et al. [34] focuses on high-value customers and studies the effectiveness of LoR in churn prediction, on a big data TelCo data set. It points to a number of questions, but focuses on the identification of influencing factors.

The authors consider the top 20% of high-value customers in the data set: a high-value customer is defined as one with a monthly expense higher than 10\$. The churn factors can be price, personal, service, product, market, marketing strategy, and market intervention of competitors. The correct discovery of these factors is the key both to recovering churned customers and to reducing the churn rate. In particular, they take 11,255 random samples (in the middle of 2020). They find a positive correlation between “Average Revenue generated Per User per month” (ARPU) and customer churn, complaints and customer churn, and a negative correlation exists between discharge of usage (DOU), current package value (pack-type), convergence business (contract), and customer churn. The highest correlation, that is, 0.5, is expressed by the current package value of the variables and ARPU.

In the article by Dalli et al. [35], the authors study: the more traditional Machine Learning techniques (NB, LoR, RF, and KNN, and SVM), some ensemble learning techniques [36] (AdaBoost, bagging, and random subspace) and DL techniques [37].

They notice that while the usage of ANN has been conducted in depth, little investigation has been done on the impact of different hyperparameters selection and explore how the hyperparameters’ tuning affects the performance of DL models following a procedure already applied in [38].

They consider an open-source database “Crowd Analytics” [39]. The data set contains 21 features and 3333 observations. The Churn rate is 14.5%. They specifically investigated the impact of the different choices for activation functions, batch size, and optimizers. The best prediction accuracy (86.8%) is derived from the combination of ReLU and sigmoid activation functions. The highest precision (84.52%) is achieved when the batch size ranges from 3 to 40 and decreases for higher values.

The best performance (86.45%) is obtained using the RMSProp optimizer. Finally, they suggest using the proposed framework to predict also the loyalty level of the customer (i.e. not loyal, loyal, or very loyal).

In the study by Melian et al. [40] they built a prediction model using a data set from a major telecommunication company in Romania. The data set consists of a sample of 10,715 postpaid customers (out of 5 million active subscribers of an anonymized database containing historical data). The number of churners is 1468 (13.70% of the sample). It contains demographic data, information about the customer’s lifecycle, financial strength, and interactions with competing TelCo clients. K-means clustering is applied to group customers into three clusters, and then several algorithms are used: LoR, DT, RF, balanced RF, and PSM. The 1505 customers which have not previously been contacted for 1 year by the TelCo service provider are included in the control group. Using LoR, DT and RF they identify the characteristics that influence most the customer’s defection:

the time span the company’s services, the number of months since changing the last offer (MonthsO), the number of minutes consumed outside the company (MinC), the value of the invoice paid for the services used, the minutes of calls received from outside the network (MinR), and the value of the extra costs paid for off-network services (ExtraCosts). The balanced RF finds that the leading variables in churners identification are MonthsO and MinC. Instead, PSM points to MonthsO, invoice, and tenure, as the three indicators that explain the net effect on the churn action of the applied treatment: “Customers have not been contacted for 12 months”. Furthermore, the 12-month contact policy—the treatment that some individuals have undergone—has affected the customers’ decision to leave the network. They conclude that MonthsO, invoice, and tenure are the main indicators that explain the net effect produced by the applied treatment on the churn action.

The article by Mustafa et al. [41] investigates which variables in the NPS, (which is a measure of customer satisfaction and loyalty using a 10-point scale) influence directly or indirectly the customer churn. Customer churn is high for customers with a low NPS. They based their study on an NPS data set from a Malaysian TelCo company (gathered in September 2019 and September 2020) consisting of 7776 records with 30 fields. They use demographic, transactional data, and NPS. They develop a propensity model for customer churn comparing the following algorithms: LR, LDA, KNN, CART, NB, and SVM. An original characteristic of this study is that the authors perform **causal mediation analysis**. They identify the **causing factors** (treated as independent variables) that result in customer churn (the dependent variable representing the effect), but also mediator variables: churn predictors can impact customer churn not only directly, but also indirectly. Another original trait of their analysis is the categorization of the customer based on the NPS value into three classes: detractor (in the range [0, 6]), passive (in the range [7, 8]), promoter (in the range [9, 10]), and the definition of “partial defection”, corresponding to a downward change of class of the customer. They find that the NPS feedback rating change partially mediates the effect of customer churn of the variables duration, reply shift, service request type, helpdesk staff ID, and assigned officer. In terms of prediction, the authors find that the CART algorithm outperforms the other algorithms tested, achieving an accuracy of 98%.

The article by Zhang et al. [42] focuses on producing a suitable customer segmentation to improve churn predictions based on machine learning (LDA and LoR) and investigate statistical methods such as Factor Analysis to identify the most significant factors in customer description. The data they use are collected from 2007 to 2018 from the three main Chinese TelCo companies (China Mobile, China Unicom and China Telecom), and include information about 4126

customers [184 women (28.7%) and 2942 men (71.3%), aged up to 107, the most common ages ranging from 20 to 60 years]. They use demographic information and business-related information, and data on customer call traffic and SMS/MMS activity.

The hypothesis behind customer segmentation is that customers with similar consumption-expense behaviors have a similar propensity to churn. Thus, the main criterion for segmentation is the expense-related characteristics. Starting from an initial set of quantities, they apply the first common factor analysis, which consists of seeking the fewest factors that can account for the common variance of a set of correlated variables.

Using LR, they find a collection of independent variables: non-monthly fixed cost; monthly fixed cost; the calls minutes of usage; long-distance and roaming call; SMS; China Unicom's MMS. Kaiser–Meyer–Olkin and Bartlett Sphericity tests are also applied to conclude that the factors are suitable for factor analysis. Given those factors and the factor gender, the authors first apply Fisher's discriminant analysis to the dimensionally reduced data. This classification technique projects high-dimensional data onto a line and performs classification in this one-dimensional space: the projection used is one that maximizes the distance between the means of the two classes while minimizing the variance within each class. The authors then use the same data to learn a classifier based on standard LoR. According to the results, the LoR-built TelCo customer churn model has higher prediction accuracy (93.94%) and better results compared to Fisher discriminant analysis (75%).

In the paper by Amin et al. [43] they discuss the **Cross-Company Churn Prediction** (CCCP) as an alternative handling of the bottleneck in within company churn prediction (WCCP). CCCP means that one company (the target) lacks enough data and can use the data of another company (the source) to successfully predict customer churn (i.e., training data from one company and applying it to the target data of another company). For this reason, they develop a model for CCCP using **data transformation** (DT) methods, i.e. z-score, log, rank and Box–Cox. The last three methods significantly improve the performance of CCCP. The authors both validate the impact of these transformation methods in CCCP and evaluate the performance of the underlying baseline classifiers: NB, KNN, GBT, SRI, and DL using publicly available datasets (i.e., subject dataset-1 and subject dataset-2). Specifically, subject dataset-1 is used as the target and contains 2850 churners and 483 non-churners. Subject dataset-2 is used as the source and contains 15,760 non-churners and 2240 churners. The evaluation measures used are confusion matrix, probability of detection (POD), probability of false alarm (POF), AUC and G-mean (GM). The NB classifier outperforms on transformed data in terms of AUC values 0.51, 0.51, 0.513 in raw, log and Box–Cox,

respectively. The DP, KNN, and GBT classifiers outperform on average, while the SRI classifier does not show significant results in terms of the commonly used evaluation measures (i.e., POD, POF, AUC, and GM). The SRI classifiers achieve the maximum performance (i.e. AUC value of 0.541) in the single DT method (i.e. Z-score) and obtain the lowest level of performance (i.e. AUC values of 0.45, 0.44, 0.357, 0.455 in raw, log, rank and Box–Cox). In conclusion, they suggest that a company that lacks the necessary data for the learning purpose of the classifier can use the data of a mature company. In addition, methodological practices are provided to assess and link the significant advantage in existing and future data transformation methods in the context of telecommunication companies. A comment is made on the data requirements that need to be considered from the researcher's perspective, as the cross-firm data should be managed in an appropriate way for empirical analysis and the development of novel models. The proposed approach provides a clear picture of the expected scientific consequences of increased data normality in terms of increased predictive model performance.

The paper by Amin et al. [44] proposes a novel adaptive customer churn prediction (ACCP) model with the ability to learn continuously (i.e., brain-like improving knowledge boundary). The machine learning at prediction time without retraining the CCP model several times with minimal data loss. In particular, an adaptive learning approach by using the NB classifier with a genetic algorithm (subclass of an evolutionary algorithm) based feature weighting approach (i.e. FWAGA: features weight assignment using GA). They employ the self-learned optimum attribute weighting technique using a genetic algorithm without losing data and keeping the attribute independence. Furthermore, the model maintains good prediction accuracy. Additionally, the authors perform a literature review on the prediction of customer churn. The ACCP performance is evaluated on publicly available data sets, considered as a benchmark, BigML Telco churn, IBM Telco, and Cell2Cell, justifying that private real-world telecommunication data sets prevent reproducibility and extrapolation. The ACCP is compared to the baseline classifiers: NB with default setting, Deep-BP-ANN, CNN, NN, LR, XGBoost, KNN, Logit Boost, SVM, and PCALB. They use accuracy, precision, recall, MCC. Achieving the results of average precision, for each data set: 0.97, 0.97, 0.98, a recall rate of 0.84, 0.94, 0.97, and F1-score of 0.89, 0.96, 0.97, an MCC of 0.89, 0.96, 0.97, and accuracy 0.95, 0.97, 0.98 respectively. They conclude that the overall performance of the ACCP approach is better by 30% in terms of average precision compared to the baseline classifier with default setting.

The Table 3 represents the information related to telephony. In particular, the fixed, mobile, mixed (i.e. fixed and mobile), unspecified (i.e. if fixed or mobile are not indicated)

Table 3 Information of fixed, mobile and mixed (fixed and mobile) telephony according to the corresponding data set that is used in each paper

Telephony	Author
Fixed (F)	[11, 51]
Mobile (Mo)	[13, 14, 19, 40, 45, 50]
Mixed (Mi)	[18, 23, 31, 42]
Unspecified (if F or Mo)	[1, 3, 9, 12, 25, 33, 35, 43, 44, 46, 52, 53]
Other fields	[24, 29, 54–57]

and other fields (i.e. if the analysis are performed in other domains) according to the corresponding data set that is used in each paper.

Discussion and Conclusion

The emergence of data services has transformed the way individuals communicate, work, and access information. The transition from conventional voice-centric communication to data-driven applications and services, such as internet browsing, social media, streaming, and online gaming, has significantly altered the landscape of the telecommunications industry. The exponential growth in mobile data consumption has impacted several factors, including network infrastructure, user behavior, and business models. It is also important to consider the role of data analytics and machine learning in optimizing network performance, personalizing user experiences, and informing strategic decision-making by mobile operators.

In this work, we surveyed the literature on machine learning methods for CCP within the scope of the TelCo domain in the approximate period of the last decade.

A remarkable fact is that, although several articles address this problem, in most cases they do not build on the findings of one another: the diversity in the structure of the data sets (e.g., presence or absence of call center contact data, presence or absence of details about the economical offer, etc.) makes the designation of a “best algorithm” impossible. In the few cases where the same public data set (referring to the same business reality) has been used, the variety of algorithms, variants, and hyperparameter configurations has hindered a definite answer to the question of the fitness of the algorithms to the problem. Definitely, the best algorithm is data set dependent. Even considering the data set of a single company over a sufficiently long time span defies the ‘best algorithm search’: the changing landscape of TelCo services makes the search for the best algorithm a rapidly moving target. Thus, the answer to the question about the best algorithm is time-dependent, company dependent, and

view-dependent (i.e. depends on the subset of company data considered).

However, there are a number of lessons learned from the literature and a number of gaps.

The definition of churner is not always unambiguous: in fixed line services, there is a clear mark of customer departure, whereas in prepaid mobile services, this is not the case: there, the definitions distinguish from one another in the number of months passed since the customer used the service the last time. Also in the fixed line services there are ambiguities: cases have been reported of some companies keeping on their books customers who had already passed to other providers and accepting to officially close the contract after complaints had been filed. This could cause distortions in the churn prediction: one would observe an intensification of the communications just before the official departure, and an algorithm could pick it up as a marker for predicting the churn, whereas in reality the customer has already left the company and is just complaining with it. In addition, many studies achieve surprisingly good results after setting apparently reasonable definitions of churner: however, in the literature there appears to be no sensitivity study checking whether they adopted exceedingly conservative definitions of the churner.

Another point refers to preprocessing issues and techniques, since prediction algorithms are just the tip of the iceberg. Leaving aside common data preprocessing problems, a variety of data preparation approaches have shown that “there is a lot of room at the bottom”: finding a good representation of the data before applying statistical learning algorithms pays off in several cases. Feature selection and feature extraction techniques (almost all articles present their own) providing this representation (or DL algorithms [32] learning this representation up front) can considerably improve the outcomes.

More strategies are available to improve prediction and target business value without trying to select the best algorithm. Focusing on value customers appears to be a sensible one, also because the group of those customers [34] seems easier to characterize for the purpose of prediction. However, the aspects of the problem left non-investigated from the business point of view (in the publicly accessible literature) are many: this is most often due to the lack of appropriate data sets, gathering the relevant information. A critical element in the customer dilemma “stay or go” is the structure and pricing of the service offer, together with possible service bundle. This kind of data is rarely publicly available. Even within large companies, gathering the different data in a single place is a challenge. Furthermore, the customer chooses also on the basis of the competing offers and finding the contextual data tagged with the appropriate timing information in the dynamic market today is almost impossible.

From a technical point of view, a notable gap in the literature is represented by the almost lack of attempts to address the problem through causal reasoning. Only the work [41] attempts a mediation analysis approach. However, to do so, it does not use the structural causal models and causal inference, but relies on correlation-based studies involving a ready available variable, the NPS, which captures domain expert knowledge on the customers.

In summary, despite the fact that the problem has a long history and that several works address its multifaceted issues, there is still much room for investigation. We have to consider the importance of the data in mobile telephony has been a **game changer**. In particular, data-driven innovations have transformed the mobile landscape, enabling unprecedented levels of connectivity, productivity, and digital engagement. Furthermore, it is important to consider the opportunities and challenges associated with harnessing the power of data in shaping the future of mobile telecommunications. The continued advancements in data technologies, coupled with effective regulatory frameworks and industry collaborations, will be essential to realize the full potential of mobile data and driving further innovation in the years to come.

Acknowledgements The work was partially supported by the Next-GenerationEU program through the project MUSA (National Recovery and Resilience Plan Mission 4 Component 2 Investment Line 1.5 (CUP G43C22001370007, Code ECS00000037), and by Italian University and Research Ministry, within the Excellence Departments program (law 232/2016) through the ReGAIInS project. The author A.B. acknowledges support from TIM S.p.A. through the Ph.D. scholarship.

Author Contributions conceptualization A.B., G.G.; funding acquisition G.G., E.D.; investigation A.B., methodology A.B., G.G., J.L., C.M., supervision G.G., F.T., E.D., A.S., writing-original draft A.B., G.G., writing-review and editing A.B, J.L., C.M., H.B., A.S. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. The work was partially supported by the NextGenerationEU program through the project MUSA (National Recovery and Resilience Plan Mission 4 Component 2 Investment Line 1.5 (CUPG43C22001370007, CodeECS00000037), and by Italian University and Research Ministry, within the Excellence Departments program (law 232/2016) through the ReGAIInS project. The author A.B. acknowledges support from TIM S.p.A. through the Ph.D. scholarship.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Vafeiadis T, Diamantaras KI, Sarigiannidis G, Chatzisavvas KC. A comparison of machine learning techniques for customer churn prediction. *Simul Model Pract Theory*. 2015;55:1–9.
2. Xia GE, Jin WD. Model of customer churn prediction on support vector machine. *Syst Eng Theory Pract*. 2008;28(1):71–7.
3. Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Huang K. Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*. 2017;237:242–54.
4. Kamalraj N, Malathi A. A survey on churn prediction techniques in communication sector. *Int J Comput Appl*. 2013;64(5):39–42.
5. García DL, Nebot A, Vellido A. Intelligent data analysis approaches to churn as a business problem: a survey. *Knowl Inf Syst*. 2017;51(3):719–74.
6. Geiler L, Affeldt S, Nadif M. A survey on machine learning methods for churn prediction. *Int J Data Sci Anal*. 2022;14:1–26.
7. Hashmi N, Butt NA, Iqbal M. Customer churn prediction in telecommunication a decade review and classification. *Int J Comput Sci Issues (IJCSI)*. 2013;10(5):271.
8. Tsai CF, Lu YH. Customer churn prediction by hybrid neural networks. *Expert Syst Appl*. 2009;36(10):12547–53.
9. Umayaparvathi V, Iyakutti K. Applications of data mining techniques in telecom churn prediction. *Int J Comput Appl*. 2012;42(20):5–9.
10. Qureshi SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management (ICDIM 2013). IEEE; 2013. p. 131–136.
11. Kirui C, Hong L, Cheruiyot W, Kirui H. Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. *Int J Comput Sci Issues (IJCSI)*. 2013;10(2 Part 1):165.
12. Brandusoiu I, Todorean G. Churn prediction in the telecommunication sector using support vector machines. *Ann Oradea Univ*. 2013;1:19.
13. Dahiya K, Bhatia S. Customer churn analysis in telecom industry. In: 2015 4th International conference on reliability, infocom technologies and optimization (ICRITO)(trends and future directions);2015. pp. 1–6.
14. Brândușoiu I, Todorean G, Beleiu H. Methods for churn prediction in the pre-paid mobile telecommunications industry. In: 2016

- International conference on communications (COMM); 2016. pp. 97–100.
15. Cao LJ, Keerthi SS, Ong CJ, Zhang JQ, Periyathamby U, Fu XJ, Lee HP. Parallel sequential minimal optimization for the training of support vector machines. *IEEE Trans Neural Netw.* 2006;17(4):1039–49.
 16. Dong J-X, Krzyzak A, Suen CY. Fast SVM training algorithm with decomposition on very large data sets. *IEEE Trans Pattern Anal Mach Intell.* 2005;27(4):603–18.
 17. Idris A, Iftikhar A, ur Rehman Z. Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Clust Comput.* 2019;22(3):7241–55.
 18. Idris A, Rizwan M, Khan A. Churn prediction in telecom using random forest and PSO based data balancing in combination with various feature selection strategies. *Comput Electr Eng.* 2012;38(6):1808–19.
 19. Ahmad AK, Jafar A, Aljoumaa K. Customer churn prediction in telecom using machine learning in big data platform. *J Big Data.* 2019;6(1):1–24.
 20. Page L, Brin S, Motwani R, Winograd T. The pagerank citation ranking: bringing order to the web, stanford digital library technologies. In: Technical report, working paper 1999-0120, Stanford Univ., Palo Alto, Calif; 1998.
 21. Kiss C, Bichler M. Identification of influencers-measuring influence in customer networks. *Decis Support Syst.* 2008;46(1):233–53.
 22. Andrews R, Zacharias R, Antony S, James MM. Churn prediction in telecom sector using machine learning. *Int J Inf.* 2019;8(2):132.
 23. ApurvaSree G, Ashika S, Karthi S, Sathesh V, Shankar M, Pamina J. Churn prediction in telecom using classification algorithms. *Int J Sci Res Eng Dev.* 2019;5:19–28.
 24. Shah M, Adiga D, Bhat S, Vyeth V. Prediction and causality analysis of churn using deep learning. *Comput Sci Inf Technol.* 2019;9(13):153–65.
 25. Śniegula A, Poniszewska-Marañda A, Popović M. Study of machine learning methods for customer churn prediction in telecommunication company. In: Proceedings of the 21st international conference on information integration and web-based applications services; 2019. pp. 640–644.
 26. Amin A, Al-Obeidat F, Shah B, Tae MA, Khan C, Durrani HUR, Anwar S. Just-in-time customer churn prediction in the telecommunication sector. *J Supercomput.* 2020;76(6):3924–48.
 27. Joolfoo MB, Jugumauth RA, Joolfoo KM. A systematic review of algorithms applied for telecom churn prediction. In: 2020 3rd International conference on emerging trends in electrical, electronic and communications engineering (ELECOM). IEEE; 2020. pp. 136–140.
 28. Yaseen A. Next-wave of e-commerce: mobile customers churn prediction using machine learning. *Lahore Garrison Univ Res J Comput Sci Inf Technol.* 2021;5(2):62–72.
 29. Bauer J, Jannach D. Improved customer lifetime value prediction with sequence-to-sequence learning and feature-based models. *ACM Trans Knowl Discov Data (TKDD).* 2021;15(5):1–37.
 30. Fader PS, Hardie BG. Forecasting repeat sales at CDNOW: a case study. *Interfaces.* 2001;31(3-supplement):94–107.
 31. Gu Y, Palaoag TD, Cruz JSD. Comparison of main algorithms in big data analysis of telecom customer retention. *IOP Conf Ser Mater Sci Eng.* 2021;1077: 012045.
 32. Jain H, Khunteta A, Shrivastav SP. Telecom churn prediction using seven machine learning experiments integrating features engineering and normalization; 2021.
 33. Lalwani P, Mishra MK, Chadha JS, Sethi P. Customer churn prediction system: a machine learning approach. *Computing.* 2022;104(2):271–94.
 34. Zhao M, Zeng Q, Chang M, Tong Q, Su J. A prediction model of customer churn considering customer value: an empirical research of telecom industry in china. *Discret Dynam Nat Soc.* 2021;2021:1.
 35. Dalli A. Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector. *Math Probl Eng.* 2022;2022:1.
 36. Gajowniczek K, Zkabkowski T, Orłowski A. Comparison of decision trees with rényi and tsallis entropy applied for imbalanced churn dataset. In: 2015 Federated conference on computer science and information systems (FedCSIS). IEEE; 2015. pp. 39–44.
 37. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–44.
 38. Domingos E, Ojeme B, Daramola O. Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. *Computation.* 2021;9(3):34.
 39. CrowdANALYTIX: Why customer churn - CrowdANALYTIX. <https://www.crowdanalytix.com/contests/why-customer-churn>
 40. Melian DM, Dumitrache A, Stancu S, Nastu A. Customer churn prediction in telecommunication industry. A data analysis techniques approach. *Postmod Open.* 2022;13(1 Sup 1):78–104.
 41. Mustafa N, Ling LS, Razak SFA. Customer churn prediction for telecommunication industry: a Malaysian case study. *F1000Research.* 2021;10:1274.
 42. Zhang T, Moro S, Ramos RF. A data-driven approach to improve customer churn prediction based on telecom customer segmentation. *Future Internet.* 2022;14(3):94.
 43. Amin A, Shah B, Khattak AM, Moreira FJL, Ali G, Rocha A, Anwar S. Cross-company customer churn prediction in telecommunication: a comparison of data transformation methods. *Int J Inf Manag.* 2019;46:304–19.
 44. Amin A, Adnan A, Anwar S. An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and naïve bayes. *Appl Soft Comput.* 2023;137:110103.
 45. Qureshi SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management (ICDIM 2013); 2013. pp. 131–136.
 46. Joolfoo MB, Jugumauth RA, Joolfoo KM. A systematic review of algorithms applied for telecom churn prediction. In: 2020 3rd International conference on emerging trends in electrical, electronic and communications engineering (ELECOM); 2020. pp. 136–140.
 47. Ahmad AK, Jafar A, Aljoumaa K. Customer churn prediction in telecom using machine learning in big data platform. *J Big Data.* 2019;6(1):1–24.
 48. Kavitha V, Kumar GH, Kumar GH, Harish M. Churn prediction of customer in telecom industry using machine learning algorithms. *Int J Eng Res Technol (IJERT).* 2020;9(5):181–4.
 49. Apurva Sree G, Ashika S, Karthi S, Sathesh V, Shankar M, Pamina J. Churn prediction in telecom using classification algorithms. *Int J Sci Res Eng Dev.* 2019;5:19–28.
 50. Jain H, Yadav G, Manoov R. Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques. In: Patnaik S, Yang XS, Sethi I, editors. *Advances in machine learning and computational intelligence.* Singapore: Springer; 2021. p. 137–56.
 51. Xu H, et al. Analysis and comparison of forecasting algorithms for telecom customer churn. *J Phys Conf Ser.* 2021;1881: 032061.
 52. Mahalakshmi S, Hemalatha M. Customer churns prediction in telecom using adaptive logitboost vs peergrading regression learning technique. *IJITEE.* 2020;9(6):1025–37.

53. Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*. 2019;7:60134–49.
54. Ghanem N, Leitner S, Jannach D. Balancing consumer and business value of recommender systems: A simulation-based analysis; 2022. arXiv preprint [arXiv:2203.05952](https://arxiv.org/abs/2203.05952)
55. Yaseen A. Next-wave of e-commerce: Mobile customers churn prediction using machine learning. *Lahore Garrison Univ Res J Comput Sci Inf Technol*. 2021;5(2):62–72.
56. Safitri AR, Muslim MA. Improved accuracy of naive Bayes classifier for determination of customer churn uses smote and genetic algorithms. *J Soft Comput Explor*. 2020;1(1):70–5.
57. Imron MA, Prasetyo B. Improving algorithm accuracy k-nearest neighbor using z-score normalization and particle swarm optimization to predict customer churn. *J Soft Comput Explor*. 2020;1(1):56–62.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.