

Aalborg Universitet



Customer retention and churn prediction in the telecommunication industry

a case study on a Danish university

Saleh, Sarkaft; Saha, Subrata

Published in:
SN Applied Sciences

DOI (link to publication from Publisher):
[10.1007/s42452-023-05389-6](https://doi.org/10.1007/s42452-023-05389-6)

Creative Commons License
CC BY 4.0

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Saleh, S., & Saha, S. (2023). Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university. *SN Applied Sciences*, 5(7), Article 173. <https://doi.org/10.1007/s42452-023-05389-6>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Research

Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university



Sarkaft Saleh¹ · Subrata Saha¹

Received: 27 February 2023 / Accepted: 16 May 2023

Published online: 03 June 2023

© The Author(s) 2023 [OPEN](#)

Abstract

In this study, we explore the possible factors affecting churn in the Danish telecommunication industry and how those factors connect with retention strategies. The Danish telecommunication industry is experiencing a saturated market regarding the number of customers, but the number of service providers has increased significantly in recent years. Due to the high costs of acquiring new customers, the telecommunication industry put great emphasis on retaining customers in such an intensely competitive industry. We employ five machine learning algorithms: random forest, AdaBoost, logistic regression, extreme gradient boosting classifier, and decision tree classifier on four datasets from two geographical regions, Denmark and the USA. The first three datasets are from online repositories, and the last one contains responses from 311 students from Aalborg University collected through a survey. We identify key features extracted by the best-performing algorithms based on five performance measures. Based on that, we aggregate all the features that appear important for each dataset. The results demonstrate that customers' preferences are not aligned. Among the prominent drivers, we find that service quality, customer satisfaction, offering subscription plan upgrades, and network coverage are unique to the Danish student population. Telecommunication companies need to integrate the sociohistoric milieu of the Nordic countries to tailor their retention policies to different consumer cultures.

Article highlights

- (i) Five machine learning algorithms were used on four datasets to extract the key factors reflecting the preferences of customers in two regions.
- (ii) Unique churn prediction and customer retention strategies are necessary for each region.
- (iii) Network coverage, customer satisfaction, service quality and subscription upgrades affect Danish students' churn.

Keywords Customer retention · Churn prediction · Classification algorithms · Telecommunication.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42452-023-05389-6>.

✉ Subrata Saha, saha@mp.aau.dk; Sarkaft Saleh, sarkaftsaleh@hotmail.com; ssaleh18@student.aau.dk | ¹Department of Materials and Production, Aalborg University, 9220 Aalborg East, Denmark.



SN Applied Sciences

(2023) 5:173

| <https://doi.org/10.1007/s42452-023-05389-6>

SN Applied Sciences
A **SPRINGER NATURE** journal

1 Introduction

Customer churn is a common problem across businesses in numerous industries, including finance [1], news [2], insurance [3], online mobile gaming [4], telecommunication [5], and online gambling [6]. According to [7], *churn management is the concept of identifying those customers who intend to move their custom to a competing service provider*. Customers might stop using a product or service for different reasons—some that might be inevitable, and others might not. Therefore, predicting which customers are likely to churn and the corresponding factors associated with their preferences are crucial to protect recurring revenue, enhance customer retention, and ensure growth [8, 9].

In the telecommunication industry, research on customer churn analysis has become increasingly important for identifying key factors affecting consumers [9]. Getting a new customer can cost five to twenty-five times more than keeping an existing customer [10]. Therefore, the industry has paid much attention to attracting new customers and retaining existing customers [11]. Empirical research also supports this; for instance, Bain & Company reported that increasing customer retention rates by 5% can increase profits from 25 to 95% [12]. Globalization and access to communication technology have resulted in multiple alternatives, motivating customers to switch from one service provider to another [13]. Due to high competition, the effects of regulatory pricing intervention and consolidation [14], and the growing popularity of over-the-top (OTT) services, the telecommunication industry faces shrinking revenues and profitability. Accurate prediction assists managers in identifying necessary actions to be incorporated into their customer relationship management (CRM), such as whether to improve the service experience, design proactive campaigns to boost adoption, or re-engage at-risk customers. Over the years, the churn prediction problem has been studied from different perspectives, such as defining churn in different contexts [15], identifying key factors affecting customers [16], developing new algorithms to improve performance accuracy [17], and reducing expenses for win-back campaigns [18].

Predicting churn manually by considering many different aspects is not straightforward, and researchers use AI solutions to automate and scale churn prediction. Churn prediction models are developed to detect customers with a high propensity to attrite [19, 20]. Many supervised machine learning algorithms (MLAs), convolutional neural network (CNN) models, and text-mining algorithms are employed to analyse churn in the telecommunication sector [5, 17]. Depending on the

features processed in the algorithms, researchers group the models into either static-based or dynamic-based [15]. By obtaining static features representative of the customer of a telecommunication service, we analyse the static behaviour of the customers.

With world-leading purchasing power, the Nordics are among the most attractive regions for telecommunication service providers. More than 26 service providers have a strong presence due to a higher penetration rate in Denmark. However, in recent years, the total revenue in this industry has remained stagnant. Evidence shows that some companies are either merged with others or acquired by others [21], (e.g., TDC acquired Plenti in 2017 and eesy A/S in 2020 [22]). Therefore, we aim to explore the following research questions: *What are the key factors affecting customer churn in the Danish student community? Which aspect(s) of the Danish telecommunication industry need more prioritization to prevent customer churn?* We used four datasets for the evaluation: three are from online repositories, namely IBM Telco, Maven Telco, and Cell2Cell; and a new dataset was collected through questionnaires from a Danish university. Five MLAs are used, and the best-performing MLAs are identified based on five performance measures for each dataset. The key features are aggregated to find similarities and distinctions between the two geographical regions. The key contributions of the present study are as follows: first, consumer preferences in the Danish telecommunication industry (DTI) have not yet been explored [9]. Therefore, this study provides insights into how customer retention policy might differ in the Nordic region. The feature of discriminating preferences can help companies to build strategic planning. Moreover, the dataset containing opinions from the Danish student population could help as a reference for further enhancement in designing churn prediction theory. Second, we identify some of the retention strategies implemented by Danish telecommunication companies. We supplement with the factors such as subscription plan offers, service quality, customer satisfaction, and network coverage to retain consumers' needs to be emphasized more in the context of Denmark.

The remainder of the study is organized as follows: In the following subsection, we present literature review and overview of the DTI. Section 2 describes four datasets and the methodologies used in this study. Section 3 presents all the results regarding MLA performance and the key features affecting churn. In Sect. 4, managerial insights are presented regarding the key difference between the DTI and other regions and how those relate to CRM. In Sect. 5, final remarks and future extension areas of the proposed study are presented.

1.1 Customer retention strategy and churn analysis

According to [23], *CRM is the strategic process of selecting customers that a firm can most profitably serve and shaping interactions between a company and these customers. The ultimate goal is to optimize the current and future value of customers for the company.* The key idea is that strategic decisions ensure improved customer satisfaction. In return, such measures result in an increase in customer retention and maintain stable growth. The literature on CRM emphasizes the use of relationship marketing strategies to identify and understand customer needs [10, 24]. Therefore, it is crucial to understand the existing customers and their needs and expectations through their purchasing patterns and sensitivity to price and loyalty programs. Studies report that a list of different activities affects and improves customer retention in the telecommunication industry, such as (i) loyalty activities [11, 25], (ii) customer satisfaction activities [26, 27], and (iii) service quality [28, 29].

Churn analysis and CRM are two sides of the same coin: the former determines the factors that cause people to leave, and the latter focuses on increasing the overall value of its customer base [30, 31]. In regard to churn analysis, Verbeke et al., [19] analyses the churn problem by applying data mining techniques (ALBA & AntMiner+) to generate accurate and comprehensible rule sets instead of a particular threshold. Note that the key performance indicator (KPI), such as profit, is also an important issue in evaluating strategies. In this context, [32] presents a profit measure for churn models, which selects a fraction of customers to include, leading to a significant increase in profits compared to traditional statistical measures. Moreover, the findings show that oversampling does not significantly improve performance. Later, the authors introduce a new expected maximum profit (EMP) criterion, which supports businesses in selecting a model that maximizes the profits and the fraction of the customer base to include in retention campaigns [33]. The implication of social network analysis models is also tested to predict customer churn [34]. Moreover, researchers also evaluated the EMP measure to compare algorithms for addressing class imbalance in churn prediction [35]. An overview of some recent studies conducted in the telecommunications industry is listed in Table 1.

Table 1 demonstrates the selection of algorithms and that their efficiency in identifying key features is not unique [8, 43, 44]. Researchers used freely available repository datasets: IBM Telco and Cell2Cell, to evaluate the performance of MLAs (e.g., XGBC, DT) by benchmarking different performance measures such as accuracy, F1-score, precision, and others. However, the datasets used in Table 1 are mostly from a particular geographical

region. It is certain that acquiring data requires both time and expense. However, based on the data of a specific area, we cannot obtain a thorough overview of preferences for customers from other regions. Due to lifestyle, social structures, regulations, and other factors, customers' preferences might differ considerably. To our knowledge, the preferences of Danish student population are missing.

1.2 Overview of the DTI

In a recent survey, Bhattacharyya et al. [9] provide a global overview of the studies in the field of churn prediction. As reported by the authors, churn prediction problems from the perspective of the DTI is sparse. The industry in Denmark has experienced a static market regarding the number of mobile cellular subscriptions from 2015 to 2020 (Fig. 1b). However, during that period, the number of registered telecommunication companies increased, and currently, there are 26 service providers. Telenor, Telia, Hi3G, and TDC remain market leaders in terms of revenue or number of subscribers. Some small-scale, low-cost service providers also exist (e.g., GreenTel, dukaTALE). Their pricing and promotion strategies might impact the gross revenue of the whole industry. The following Fig. 1 present an overview of total revenue in the telecom industry, subscriptions, and the number of service providers in Denmark.

Therefore, it is highly challenging for Danish companies to maintain stable financial growth due to increasing competition. In this regard, predicting the key factors influencing churn is crucial for them. As shown in Fig. 1c, the DTI has experienced increasing growth in the number of service providers. After the COVID-19 pandemic, people are more familiar with OTT media services, and their subscriptions affect gross revenue. Moreover, telecommunication service providers utilise substantial fixed and shareable infrastructure that must be offset by revenue. The population growth in Denmark over the past ten years has remained stagnant [45] and acquiring new customers is a challenging issue. Identifying factors affecting churn can benefit telecommunication service providers to retain customers by minimizing customer churn. Therefore, our study can provide insights into the key factors to be emphasized for future business.

2 Methodology

2.1 Data used for the evaluation

The data used in this study consist of three online available repository datasets: IBM Telco, Cell2Cell and Maven Telco [5, 36]. A survey was conducted to collect responses of students from Aalborg University, because no online

Table 1 Literature review of churn prediction in telecommunication industry

Study	Dataset(s)	Method(s)	Evaluation metric(s)	Key remark(s)
[36]	IBM telco & Cell2Cell	Deep-BP-ANN, LR, NB, KNN & XGBC	Acc., AUC, Recall Precision & F1-Score	Deep learning method outperform existing ML methods in performance measures.
[5]	IBM telco & Cell2Cell	KNN, RF, LR, NB, GBC, XGBC, AdaBoost & ETC	Acc., Precision Recall & F1-Score	Cell2Cell: RF, NB, XGBC & AdaBoost. IBM: LR, NB & XGBC A proposed combination of classifiers outperform all in terms of accuracy.
[13]	Cell2cell & Orange	DL, NB, DT, GBC, RF	Acc., Precision, Recall & F-score	Ensemble classifiers outperforms the single classification techniques.
[17]	IBM telco	CNN	Accuracy	Proposed CNN model support businesses to predict CC by TPR of 95%.
[37]	Larose Telco, Telecom1, Cell2cell, IBM telco, Telecom2, Asian Telco	LDA, SVM, RF, Project-pursuit RF	Acc., RMSE, KS, AUC, Precision, Sens., Spec., Gini, OOB & F-score	Projection pursuit RF perform better predictive than utilised methods.
[38]	Dataset-1, IBM telco, PAKKD & Dataset-4	Bayesian binomial test Distance factor	Precision, Recall, F-measure & Acc.	Non-churn customers and churn have resembling features which negatively effects classification failure rate, therefore, focus on level of certainty is crucial to consider.
[16]	Syria-Telco	XGBC, GBM-Tree, DT, RF	AUC	XGBC outperform all algorithms presenting 93.301% AUC.
[39]	Cell2cell	DT, ANN, SVM, kNN	AUC, Acc., Sens., Spec. & F-measure,	Combining SOM with classifier ensembles by PCA provide best prediction results. Filtering out irrelevant features is important for data pre-processing.
[40]	Dataset-1	4 FUZZY classifiers & 9 NON-FUZZY classifiers	TP & FP rate, AUC	VQNN (FP rate) and OWANN (TP rate & AUC) FUZZY classifier outperform all.
[41]	UCI repository	SVM (Kernel = RBF, LIN, POL & SIG)	Acc. & Gain	SVM(POL) perform best with 88.56% accuracy.
[42]	Taiwan-Telco	DT & BP-ANN	Lift & Hit-Ratio	BP-ANN outperform DT

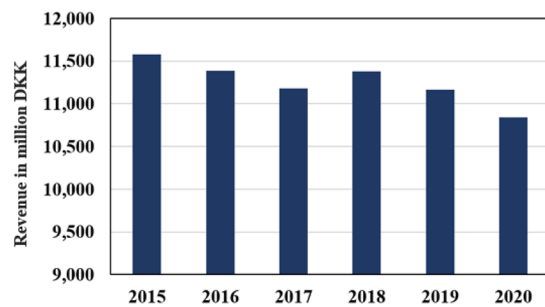
ACC accuracy, ANN artificial neural network, AUC area under the curve, BP-ANN back propagation artificial neural network, CC customer churn, CNN convolutional neural network, DL deep learning, ETC extra tree classifier, GBC gradient boosting classifier, GBM gradient boosting machine, KNN k nearest neighbor, KS kolmogorov-smirnov test, LDA linear discriminant analysis, LIN linear, LR logistic regression, NB naive bayes, OOB out-of-bag error, OWANN ordered weighted average nearest neighbour, PCA principal component analysis, POL polynomial, RBF radial basis function, RF random forest, RMSE root mean squared error, SENS sensitivity, SIG sigmoid, SOM self-organizing map, SPEC specificity, SVM support vector machine, TPR true positive rate, VQNN vaguely quantified nearest neighbors, XGBC extreme gradient boosting classifier

repository data are available in reflecting Danish customer preference in the telecommunication industry. Therefore, the survey data can reflect student preferences in this country. We designed our questionnaire to collect data on four categories: (i) customer care details, (ii) customer demography, (iii) payment methods, and (iv) value-added services. The survey was a mixed-method survey; some questions required quantitative answers and qualitative responses. Each dataset is described below:

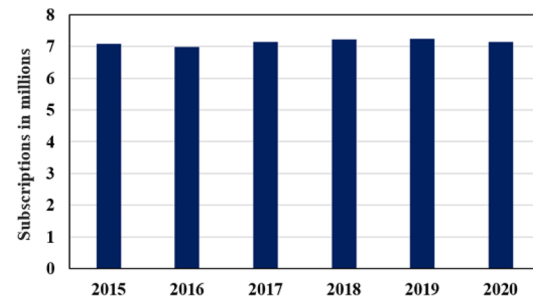
IBM Telco data are provided by International Business Machines (IBM) Corporation and extracted from the Kaggle Data Platform [46]. The dataset consists of 7043 samples with 21 attributes. In addition, it includes information

about customer churn and three numeric features, such as tenure period, the monthly and total charges from each customer. The categorical features includes for each customer information about service attributes (internet service, security, device protection, streaming TV, etc.), information about the customer account (type of contract, payment method, etc.), and demographic information about customers (gender, age, etc.). The IBM Telco dataset is imbalanced, consisting of 1869 churned customers and 5174 nonchurned customers registered in California, USA.

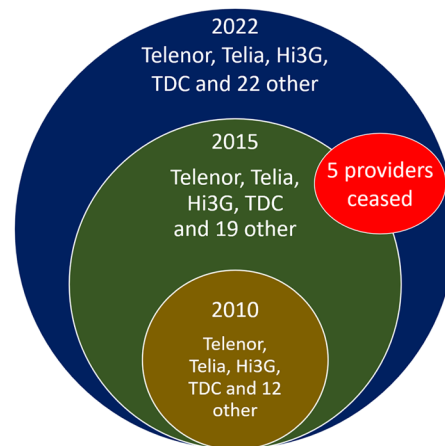
Maven Analytics Telco churn data are sourced from IBM and extracted from the Maven Analytics Data Ground [47]. The dataset is constructed with attributes similar to



(a) Revenue of cellular telephony and data in million(s) DKK, [73]



(b) Number of cellular subscriptions in million(s), [73]



(c) Telecommunication providers in Denmark, [69], [70], [72]

Fig. 1 Characteristics of the DTI

the IBM telco dataset but supplemented with additional features such as average monthly GB download, unlimited data option, age, streaming music, long distance charges, etc. The full list of features are listed in 38 columns. The dataset has been cleaned for missing values or empty features, and customer status is restricted to either *churned* or *stayed*. Similarly, with the IBM Telco dataset, samples are from customers in California, USA. Maven Telco dataset consist of 1586 churned customers and 3015 nonchurned customers.

Cell2Cell data are from the Teradata Center for CRM of Duke University USA and extracted from the Kaggle Data Platform [48]. The dataset consist of 71,047 samples and 58 attributes. The dataset contains information about customers including (i) value-added services, (ii) demographic information, (iii) payment and bill method, (iv) usage patterns. The samples are from one of the largest wireless companies in the USA. Cell2Cell data is imbalanced, which consists of 14,257 churned customers and 35,519 nonchurned customers registered in USA.

AAU data are collected through a survey carried out at Aalborg University (AAU), Denmark. A total of 311 students participated in the survey, and 288 provided complete feedback based on 21 questions (the number of questions for each respondent varied according to the answers). The authors designed the survey questionnaire based on sixteen quantitative and five qualitative questions. For instance, we acquire responses against the question: *Which of the following criteria do you find the most important while choosing a service provider?*, where the respondent could choose among one or several options: price, streaming services, data GB for roaming, online services (protection, back-up), high data GB, professional brand, call hours for roaming, promotional offers, or other. The respective office assistants sent the survey questionnaire to undergraduate/postgraduate students in Materials and Production, Energy, Architecture, Design, & Media Technology, and Politics & Society departments. The aim of involving the office assistants was to ensure anonymity and allow respondents to share their thoughts truthfully. A five-point Likert scale was used so students

Table 2 Datasets after exclusion of irrelevant features and missing values

Dataset (Version)	Region	No. of features	Sample size	No. of churned customers (%)
IBM Telco	USA	19	7043	1869 (26.5)
Maven Telco	USA	27	4601	1586 (34.5)
Cell2Cell	USA	39	49,776	14,257 (28.6)
AAU	Denmark	18	288	109 (37.8)

could choose from two extremes, two intermediates, and one neutral option. Excluding incomplete responses, the dataset consisted of 109 churned samples and 179 non-churned samples. We refer to Table 2 for an overview of the four datasets. Additionally, we refer to Tables S2.1 and S2.2 in the Supplementary file for the description of features for IBM Telco data, Maven Telco data, and Cell2Cell data.

2.2 Machine learning algorithms

It is extremely challenging to prepare a proactive retention strategy if the methodology inaccurately identifies the key factors affecting customer churn. Otherwise, it can mislead managers and require them to change their action plan. The performance measures of each MLA can be very different when implemented in different datasets for many reasons, such as hyperparameter settings, dataset segmentation, and computation power. Therefore, their feature importance might also not be aligned. The permutation feature importance technique is utilised in this study using Python by importing the Scikit-learn module; *permutation feature importance*. The technique is defined as the relationship between the feature and the target. A drop in the model accuracy indicates how much the developed model depends on the specific feature. Therefore, we rely on the outcomes of five algorithms to aggregate important features using the Python module Scikit-learn. We applied synthetic minority oversampling technique (SMOTE) to the AAU data to balance and analyse the difference [49, 50]. In real-world data, the majority of samples belong to one class (*nonchurners*), while the more important class is typically the minority group (*churners*). Previous studies have shown that applying MLAs to large imbalanced dataset tend to yield poor performance [49, 50]. The SMOTE is widely used in the field of churn prediction to overcome this challenge. It generates artificial data for the minority class for each original minority sample while not taking into account the instances from the majority class [35]. This can result in not representing the real-world characteristics of a dataset, hence, the outcome may not reflect the true nature of the minority class. We refer to Section S5 in the Supplementary file for details.

Random forest (RF) is an effective classification method with nonlinear data. A number of decision trees are created by choosing any random sample of attributes from the predictor attribute set. The final decision tree mainly uses weighted averages for the predictions. Unlike other algorithms, it performs better if correlated features exist in the data [51]. RF is utilised for churn prediction by [5, 18].

Adaboost Classifier (ADA) is an ensemble classifier that iteratively retrains the algorithm by selecting the training set based on the effectiveness of prior training. While a single algorithm might not be very effective at classifying objects, the overall classifier can achieve a high accuracy score when paired with a boosting ensemble algorithm. It enhances the algorithm's prediction performance by turning a group of weak learners into strong learners [51]. ADA is used for churn prediction by [5, 52].

Logistic regression (LR) uses the logistic function to squeeze the output rather than fitting it to a straight line or hyperplane. The function does not have a linear relationship with weights. A probability is created from the weighted sum using the logistic function. It uses a maximum-likelihood estimator to maximize the probability of the observed outcome. It computes the error for each prediction and the prediction value for each instance in the training set. This procedure repeats until the model is sufficiently accurate [51]. LR algorithm is used for churn prediction by [5, 36].

Extreme gradient boosting classifier (XGBC) is a decision tree-based ensemble technique created using gradient boosting. The objective function minimized by XGBC combines a penalty function for model complexity with a loss function. The gradient boosting method first computes the residuals of the previously applied model using new models, and then it combines both of these results. It is known as "gradient boosting" because it minimizes loss when introducing new models by employing a gradient algorithm [53]. XGBC algorithm is used by [5, 36] for churn prediction.

Decision tree (DT) is constructed as a tree by adding nodes along each branch until it reaches a terminal node. Each node in a decision tree is given a class label and represents a test on its attributes. The branches descending from each node in the DT show all potential values that an attribute may take. The advantages of decision tree algorithms are that they are quite resistant to noise and are relatively simple to interpret [51]. DT is used in recent studies for churn prediction [13, 39].

2.3 Performance measure

Five measures are evaluated for each algorithm to find the best-performing. The accuracy metric ($Accuracy = \frac{TP+TN}{TP+TN+FN+FP}$)



Fig. 2 Computational scheme for study and key features

is used to measure the total number of correctly identified instances. The precision metric ($Precision = \frac{TP}{TP+FP}$) is used to measure how the model is observing the actual number of positives against the predicted positive. The negative predictive value ($NPV = \frac{TN}{TN+FN}$) is used to measure how the model identifies the non-churners. F1-score ($F1 - score = \frac{2TP}{2TP+FP+FN}$) is used to measure how accurate the algorithm's performance is. Finally, area under the curve (AUC) is calculated by ($AUC = \frac{1}{2} - \frac{FP}{2(FP+TN)} + \frac{TP}{2(TP+FN)}$), where TP true Positive; TN true Negative; FP false Positive; and FN false Negative. A higher result indicates a more accurate performance. In addition, top-decile lift is measured for best performing MLAs for each dataset.

3 Results

Before implementing MLAs, we conduct preanalysis to exclude samples with missing values and features that are not relevant (e.g., we exclude features such as *marital status* for the Cell2Cell dataset). We refer to Section S2 in the Supplementary file for the detailed features we considered for the analysis. Each dataset is divided into two subsets: a training set and test set with a 70/30 ratio. The first subset is used to fit the algorithms, and the test set is used to predict unseen data. We refer to Table S4.1 in the Supplementary file for the details of the parameter settings used to implement each algorithm. Furthermore, the best-performing MLA(s) based on each performance measure was determined for each dataset. Figure 2 presents the computational scheme used and the results obtained in this study.

From Fig. 2, we find that the key features for each classification algorithm are not aligned but not wholly different. For clarity, we present the performance measure for each MLA in four datasets in Fig. 3.

We refer to Fig. 3 demonstrate that in terms of each performance measurement, LR (Acc., F1-score, AUC & NPV) and RF (Precision) outperformed other MLAs and are the best-performing algorithms for IBM Telco dataset. Similarly, for the Maven dataset, XGBC (Acc., F1-score, AUC & NPV) and RF (Precision) outperformed other utilised MLAs. For the Cell2Cell dataset, XGBC (Acc., AUC & NPV), RF (Acc. & precision) and DT (F1-score) outperformed other MLAs. Finally, for the AAU dataset, RF (Acc., F1-score, AUC & NPV) and XGBC (Precision) outperformed the other implemented MLAs. Furthermore, Tables S4.1 to S4.4 in the Supplementary file present the top-decile lift for the best-performing algorithms for each dataset. The top-decile lift focuses exclusively on the most critical group of customers and their churn risk. Gain shows the percentage of actual churners covered at a given decile level, whereas lift indicates the ratio percentage to the random rate at a given decile level. As shown in Table S4.3 in the Supplementary file, in decile level 4, XGBC obtains 43.8% of churners covered in the top 40% of the data. A lift of 1.1 indicates that in 40% of the data based on the model, we could most likely find the churners 1.1 times more than randomly selected 40% of the data without a model. We apply the SMOTE technique to balance the two classes of the AAU dataset. Although we observe an increase in performance in some algorithms, we obtain no changes in the best-performing MLAs based on five different performance measures (Table S5.1 in Supplementary file). Nevertheless, XGBC and RF outperformed the other MLAs, and their feature importance is presented in Table S5.2 in

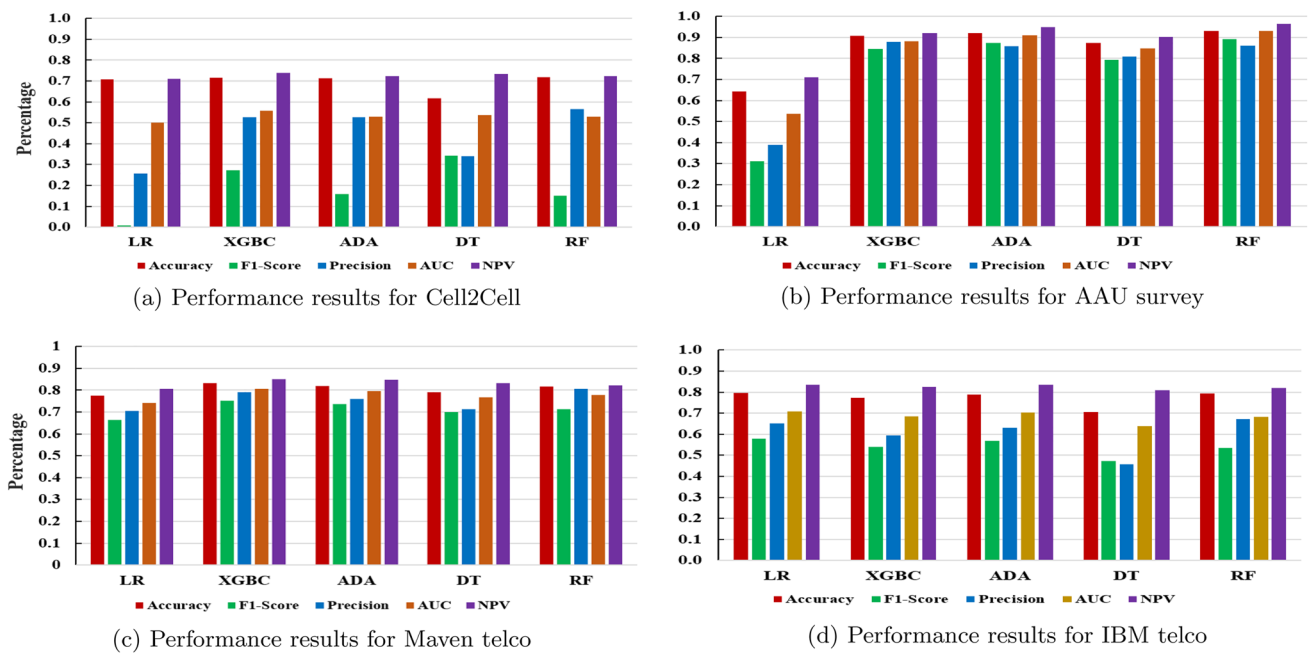


Fig. 3 Performance results of the test set based on five different measures

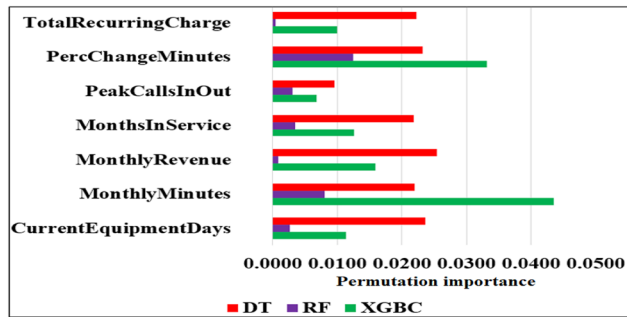
the Supplementary file. We observe mostly similar important features for the balanced/imbalanced dataset with *PaymentMethod* included and *PackageSatisfaction* and *CustomerService* excluded for the latter. As we obtain significantly high performance without data manipulation by applying data balancing techniques (see Table S3.1 in Supplementary file), we present our conclusion based on those unbalanced datasets to avoid any recommendation with synthetic data samples included.

The result supports that the key features identified by the MLAs are not aligned. For example, if we look at the important features obtained through the AdaBoost algorithm, which is also popular in churn analysis for telecommunications data [52], the key features are significantly different compared to those identified by other algorithms. Noticeably, the decision tree algorithm performs poorly compared to all four other algorithms in our study (see Table S3.1). Similarly, the LR algorithm is outperformed by the other four algorithms in the Maven, Cell2Cell, and AAU datasets. Therefore, the five algorithms' implementations help us aggregate possible features in two geographical areas. Note that we can exclude the LR algorithm if we only perform churn analysis based on three datasets. We also found that LR identified *brand* as a key feature, which was not recognized by the other algorithms. Note that, in the telecommunication industry, *brand* is also a key [54].

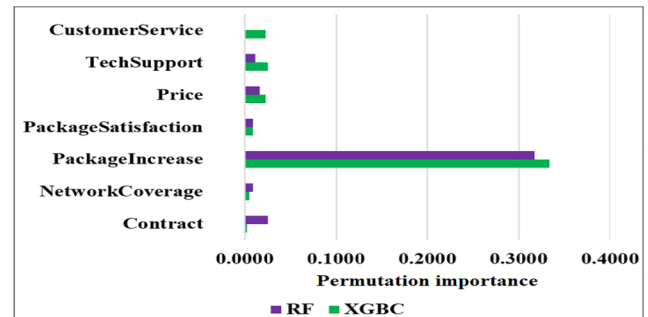
Figure 4 presents the top features affecting customers to churn for each dataset from the best-performing algorithms and their importance. In Tables S3.4–S3.7, the numerical values present all features used to predict

churners. The higher the value, the more important that specific feature is to the churn risk. The values indicate how much each algorithm depends on that particular feature to identify the churners from each dataset. For instance, Fig. 4c demonstrates that contract type and the monthly charge impact customer churn most. Table S3.3 in the Supplementary file presents each algorithm's top five features affecting customer churn.

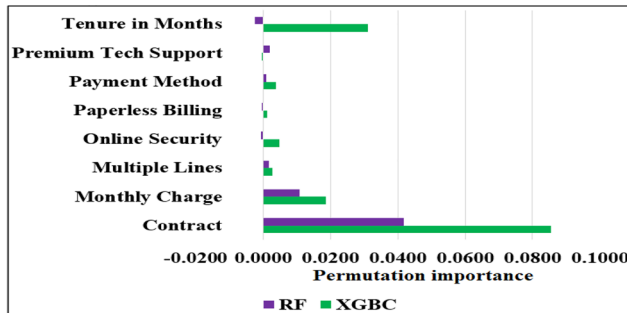
Recall that Fig. 2 presents six unique features from the IBM Telco dataset, seven from the Maven Telco dataset, seven from the Cell2Cell dataset, and seven for the AAU dataset when combining features from the best-performing MLAs. Aggregating features that appear significant in datasets from the USA, we obtain a list of 13 unique features. Some of these features appear as key in the context of USA customers but not for AAU respondents. Some features such as *Age*, *CurrentEquipmentDays*, *InternetService*, *Online Security*, *PhoneService* and *PeakCallsInOut*, are not directly related to the AAU dataset because they have not been found to be relevant or due to the lack of data availability. For instance, we did not consider age as a feature in the AAU dataset due to a focus only on undergraduates/postgraduates in the survey. Since most Danish companies provide additional data, the *InternetService* feature is not applicable to the DTI. However, in the USA, approximately 10% of mobile phone subscribers do not subscribe to wireless internet access [55]. The AAU survey includes other features such as *PhoneService* and *OnlineSecurity* (Fig. S1.2b), which were prioritized by most of the respondents. However, compared to price, data usage, and other factors,



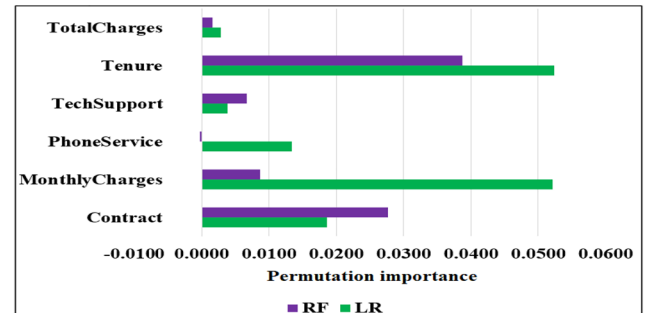
(a) Feature importance for Cell2Cell



(b) Feature importance for AAU data



(c) Feature importance for Maven Telco



(d) Feature importance for IBM Telco

Fig. 4 Feature importance from best-performing algorithms for each dataset

these features were given lower priority. A list of features appears important for the DTI but not for the USA's telecommunication industry. Due to the easy access of other EU nations, network coverage, that is, accessibility of the internet and calling in other countries, becomes key as the participants are students enrolled from different countries (refer to the question in Fig. S1.3 in Section S1 in Supplementary file). For instance, network coverage is identified as *RoamingCalls* and *non-US travel* in Cell2Cell data but does not appear as important when we see the best-performing algorithms. *PackageIncrease* feature is included in questions (refer to Fig. S1.1j), and the motivation was to verify whether the respondent had been offered any promotion by their existing provider, such as an increase in subscription plan offers or other value-offering (e.g., gift certificate, online services included). This is identified in Cell2Cell data as *CustomerCareCalls* and *RetentionOffersAccepted*, but these features do not appear as key for the USA telecommunication industry. Moreover, questions in Fig. S1.3 present customer service and package satisfaction as features for AAU data. From USA data, these features are identified as *CustomerCareCalls*, *RetentionCalls* and *RetentionOffersAccepted*, which appears important only for the DTI.

We also identified key features that appear important in the AAU data as well as aggregated features in the other three datasets. For instance, the *Contract* feature appears important

for IBM Telco and Maven Telco data, and this finding also emerges from AAU data. Similarly, we found that price features affect customer churn for Danish telecommunication, which also appears crucial for the USA industry for all datasets in terms of *MonthlyCharge*, *MonthlyMinutes* and *PercChangeMinutes*. We also found *TechSupport* to be essential in both the Danish and USA telecommunication industries.

4 Discussion

Classification results demonstrate that a single MLA fails to ensure the best performance in all five performance measures. We found that two MLAs outperform other algorithms based on one or several performance measures for a dataset. These findings are well-suited to recent studies. For instance, [5] showed that LR ensured the highest F1-score (we also found LR outperformed based on accuracy and F1-score) for the IBM Telco dataset. Similarly, for the Cell2Cell dataset, [5] found XGBC to perform best in terms of accuracy and precision metrics (for accuracy metrics, we found XGBC and RF, and RF also ensured the highest in terms of precision). Therefore, we suggest that instead of relying on a single MLA, one should apply multiple MLAs to develop a set of important features.

We found some interesting areas: First, international network coverage appears important for students, but

the number of respondents was too low. Since students are from a different region in the EU, outside calling facilities and internet usage are also key. This particular feature appears important, as shown in Fig. S1.2b and S1.3. Similarly, we found that five respondents switched as their parents switched. For them, parents were responsible for their subscriptions. In Europe, a parent sometimes pays the student's subscription [56].

Analysing churn and initiating customer retention strategies has been a research priority in recent years [18]. In some studies, researchers only focus on either churn analysis [13, 36] or forming customer retention strategies without detailed background [57]. However, those are not separate. Therefore, we focus on the interlink between churn prediction and customer retention. As Jeff Bezos once quoted, *"We see our customers as guests to a party, and we are the hosts. It's our job every day to make every important aspect of the customer experience a little bit better"*¹. In this study, the findings of feature importance for the telecommunication industry on two geographically different datasets show both similarities and differences. However, information from the extracted datasets contains common features, e.g., Cell2Cell data are sourced from a telecommunication provider in the USA containing data about service usage such as call details and call duration, which we did not consider in the survey. For instance, we identified some important features for the USA regarding customer service usage, such as *PeakCallsInOut* and *PercChangeMinutes*. In the telecommunication sector, customers sometimes select the service provider based on their history and reputation. It might be a possible way to analyse key features based on each service provider. However, due to the sample size, it does not appear a feasible way to analyse AAU data. Similarly, with [58], we found that customer service (*customer care calls*) is not a key factor for the Cell2Cell dataset affecting customer churn. Nevertheless, we found it important in the context of the DTI. Similar to our findings, studies by [28] and [29] show that service quality greatly influences retention in the telecommunication industry. Additionally, we found that package satisfaction is an important feature in the Danish industry; similarly, with [56], the study presents that customer satisfaction affects the intent to continue service and is also observed in the Nordic region [54]. According to recent evidence, features such as price and satisfaction affect churn and are considered by business managers in the DTI. However, we also found that telecommunication companies operating in Denmark should consider implications such as upgrades in subscription plans and network coverage, as they affect customer churn.

¹ <https://www.goodreads.com/quotes/794527-we-see-our-customers-as-invited-guests-to-a-party>.

According to [59], existing customers in the Nordic region receiving discounts can be retained and are stimulated to purchase additional services from their existing provider more than customers without any discount offers.

In contrast to global telecommunication markets, e.g., the USA, China, or India, initiating customer retention strategies is especially crucial in the DTI. Customer retention is important for businesses experiencing saturated markets or lower growth of new customers as experienced in the DTI [60, 61]. The laws and regulations of the European Union have impacted the telecommunications industry in all markets of the European nations, as stated by the CEO of Vodafone Group [62]. Only three or four major mobile network operators (MNOs) serve a larger population on other markets in comparison to over 100 MNOs competing in the European market. In addition, the laws and regulations imposed on MNOs make it possible for mobile virtual network operators (MVNOs) to operate on their cellular networks. This has enabled telecommunication companies to exist with low costs and disrupt the telecommunication industry by solely competing on price, as observed in the DTI. Moreover, the laws and regulations of the EU prohibit major companies from merging, e.g., the merger of Telenor and Telia in 2015 was rejected to prevent them from forming a dominant company in the DTI [63].

In some recent initiatives, the major telecommunication companies in Denmark have shifted away from price-competing towards promoting campaigns with service attributes included in the subscriptions. For instance, Telia recently promoted a new brand strategy as one of the most expensive campaigns to offer cellular services, TV streaming, and internet as package deals targeted at families [64]. Telenor has recently shifted its focus away from highly competitive price offers to differentiate itself from the market by providing reassuring services such as safe internet browsing and screen switching. The result has been a success, presenting the best annual results in ten years and the highest quarterly satisfaction in the company's history of operating in Denmark [64–67]. Tenure appears to be an important feature for US telecommunication in all three datasets in terms of *Tenure* or *MonthsinService* but does not appear to be important for Danish telecommunication. However, the question in Fig. S1.1f shows that we grouped the tenure period in monthly intervals, which are listed as units in months for USA data.

5 Conclusion and future research

The DTI is experiencing saturation, and new strategies in addition to competing on price are necessary. Churn analysis and customer retention strategies are key in the highly competitive industry to predict churners and

initiate proactive activities to retain existing customers. Five MLAs were implemented, and the performance was evaluated based on five different measures. The best-performing algorithms identified important features affecting customer churn in the telecommunication industry in the USA and Denmark. The results suggest that the DTI should upgrade subscription plan offers to retain existing customers and focus on service quality, customer satisfaction, and network coverage. We found that age is an important factor influencing churn [68], although we restricted our focus to undergraduate/post-graduate students. In the future, heterogeneity needs to be considered.

Theories of customer churn in the telecommunication industry are most often ingrained in the USA. However, the key factors affecting consumers might differ if investigated in a wider range of cultural and socioeconomic contexts. Therefore, to supplement recent efforts to develop more localized consumer culture and retention theory [69], we conduct churn analysis between two regions to explore more macro perspectives. The contribution has the potential to formulate conceptual dialogues for the telecommunication industry in the Nordic region, which is somehow aligned with the USA's format but not utterly. We demonstrate new features that can be put on the springboard for strategy refinement. As we projected, a limitation of our survey is that we do not have evidence about the characteristics of the retention offers to Danish consumers and do not know whether targeted consumers accept them. The number of respondents for the AAU dataset represents the opinions of a particular age group, which might fail to provide an overview of the whole population. The number of respondents is also limited. There might be additional features, such as the effect of social networking and profitability, which can be included. The dataset also lacks call-detail records (CDRs) (such as the duration of calls or the frequency of communication between two customers). In future research, social network analytics may provide the opportunity to understand customer behaviour and further enhance churn prediction accuracy. One of the central challenges is accurately defining the social network based on available relational data, e.g., CDRs. Additionally, processing such large amounts of data from CDRs can be a difficult and time-consuming task as it often contains millions of records and requires considerable preprocessing before being able to utilise for analysis [34]. From the theoretical point of view, the static measure we used could be complemented by dynamic analysis, anticipating the possibility of proposing retention offers afterwards. However, the key features that discriminate telecommunication customers in two geographical regions can be considered a starting point for future evaluation

and supplement churn prediction theory from a global perspective.

Acknowledgments The authors are thankful to their department secretary Hanne Korgaard Skjellerup for her support in arranging the survey.

Author contributions S. Saleh and S. Saha collected data, decide inclusion and exclusion features and coordinated research. S. Saha contributed to the conception of the study. S. Saleh performed computational work. Both authors equally contributed to write the manuscript, revised the manuscript, and approved the final version.

Data availability All relevant data collected from Aalborg University are available on request from storage platform Data deposit at Aalborg University in the website: <https://doi.org/10.57957/datadeposit.1fad02be-0b86-4281-bc5a-c92527172105> and S. Saleh.

Declarations

Competing interest The authors declare that they do not have any competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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. Kaya E, Dong X, Suhara Y, Balcisoy S, Bozkaya B (2018) Behavioral attributes and financial churn prediction. *EPJ Data Sci* 7(1):41. <https://doi.org/10.1140/epjds/s13688-018-0165-5>
2. Ballings M, Van den Poel D (2012) Customer event history for churn prediction: how long is long enough? *Expert Syst Appl* 39(18):13517–13522. <https://doi.org/10.1016/j.eswa.2012.07.006>
3. Günther CC, Tvete IF, Aas K, Sandnes GI, Borgan Ø (2014) Modelling and predicting customer churn from an insurance company. *Scand Actuar J* 2014(1):58–71. <https://doi.org/10.1080/03461238.2011.636502>
4. Perišić A, Jung DŠ, Pahor M (2022) Churn in the mobile gaming field: establishing churn definitions and measuring classification similarities. *Expert Syst Appl* 191:116277. <https://doi.org/10.1016/j.eswa.2021.116277>
5. Beeharry Y, Tsokizep Fokone R (2022) Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. *Concur Comput: Pract Exp* 34(4):e6627. <https://doi.org/10.1002/cpe.6627>

6. Coussement K, De Bock KW (2013) Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning. *J Bus Res* 66(9):1629–1636. <https://doi.org/10.1016/j.jbusres.2012.12.008>
7. Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: state-of-the-art and future trends. *Comput Operat Res* 34(10):2902–2917. <https://doi.org/10.1016/j.cor.2005.11.007>
8. Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Huang K (2017) Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing* 237:242–254. <https://doi.org/10.1016/j.neucom.2016.12.009>
9. Bhattacharyya J, Dash MK (2022) What do we know about customer churn behaviour in the telecommunication industry? A bibliometric analysis of research trends, 1985–2019. *FIB Bus Rev* 11(3):280–302. <https://doi.org/10.1177/23197145211062687>
10. Singh R, Khan IA (2012) An approach to increase customer retention and loyalty in B2C world. *Int J Sci Res Publ* 2(6):1–5
11. Wong KKK (2010) Fighting churn with rate plan right-sizing: a customer retention strategy for the wireless telecommunications industry. *Serv Ind J* 30(13):2261–2271. <https://doi.org/10.1080/02642060903295669>
12. Gallo A (2014) The Value of Keeping the Right Customers. <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>. Accessed 9 Oct, 2022
13. Liu R, Ali S, Bilal SF, Sakhawat Z, Imran A, Almuhaimeed A, Sun G (2022) An intelligent hybrid scheme for customer churn prediction integrating clustering and classification algorithms. *Appl Sci* 12(18):9355. <https://doi.org/10.3390/app12189355>
14. Farooq M, Raju V (2019) Impact of over-the-top (OTT) services on the telecom companies in the era of transformative marketing. *Glob J Flex Syst Manag* 20(2):177–188. <https://doi.org/10.1007/s40171-019-00209-6>
15. Alboukaey N, Joukhadar A, Ghneim N (2020) Dynamic behavior based churn prediction in mobile telecom. *Expert Syst Appl* 162:113779. <https://doi.org/10.1016/j.eswa.2020.113779>
16. Ahmad AK, Jafar A, Aljoumaa K (2019) Customer churn prediction in telecom using machine learning in big data platform. *J Big Data* 6(1):1–24. <https://doi.org/10.1186/s40537-019-0191-6>
17. Tariq MU, Babar M, Poulin M, Khattak AS (2021) Distributed model for customer churn prediction using convolutional neural network. *J Model Manag*. <https://doi.org/10.1108/JM2-01-2021-0032>
18. Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW (2019) A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access* 7:60134–60149. <https://doi.org/10.1109/ACCESS.2019.2914999>
19. Verbeke W, Martens D, Mues C, Baesens B (2011) Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Syst Appl* 38(3):2354–2364. <https://doi.org/10.1016/j.eswa.2010.08.023>
20. Mobilabonnement. (2022). Liste over danske mobilskaber. <https://mobilabonnement.dk/mobilskaber/>. Accessed 9 Nov, 2022
21. Grajek M, Gugler K, Kretschmer T, Mişçişin I (2019) Static or dynamic efficiency: horizontal merger effects in the wireless telecommunications industry. *Rev Ind Organ* 55(3):375–402. <https://doi.org/10.1007/s11151-019-09723-4>
22. Olsen L (2021) Overblik over de danske teleselskaber - hvilke teleselskaber ejer hvem?. <https://24tech.dk/nyheder/telekommunikation/teleselskaber-hvem-ejer-hvem/>. Accessed 17 Nov, 2022
23. Kumar V, Reinartz W (2018) Customer relationship management, 3rd edn. Springer-Verlag GmbH, Berlin. <https://doi.org/10.1007/978-3-662-55381-7>
24. Sigala M (2005) Integrating customer relationship management in hotel operations: managerial and operational implications. *Int J Hosp Manag* 24(3):391–413. <https://doi.org/10.1016/j.ijhm.2004.08.008>
25. Lam SY, Shankar V, Erramilli MK, Murthy B (2004) Customer value, satisfaction, loyalty, and switching costs: an illustration from a business-to-business service context. *J Acad Mark Sci* 32(3):293–311. <https://doi.org/10.1177/0092070304263330>
26. Kim MK, Wong SF, Chang Y, Park JH (2016) Determinants of customer loyalty in the Korean smartphone market: moderating effects of usage characteristics. *Telematics Inform* 33(4):936–949. <https://doi.org/10.1016/j.tele.2016.02.006>
27. Kumar V, Dalla Pozza I, Ganesh J (2013) Revisiting the satisfaction-loyalty relationship: empirical generalizations and directions for future research. *J Retail* 89(3):246–262. <https://doi.org/10.1016/j.jretai.2013.02.001>
28. Deng Z, Lu Y, Wei KK, Zhang J (2010) Understanding customer satisfaction and loyalty: an empirical study of mobile instant messages in China. *Int J Inf Manage* 30(4):289–300. <https://doi.org/10.1016/j.ijinfomgt.2009.10.001>
29. Santouridis I, Trivellas P (2010) Investigating the impact of service quality and customer satisfaction on customer loyalty in mobile telephony in Greece. *TQM J*. <https://doi.org/10.1108/17542731011035550>
30. Ascarza E, Neslin SA, Netzer O, Anderson Z, Fader PS, Gupta S, Schriff R (2018) In pursuit of enhanced customer retention management: review, key issues, and future directions. *Cust Needs Solut* 5(1):65–81. <https://doi.org/10.1007/s40547-017-0080-0>
31. Krishna GJ, Ravi V (2016) Evolutionary computing applied to customer relationship management: an survey. *Eng Appl Artif Intell* 56:30–59. <https://doi.org/10.1016/j.engappai.2016.08.012>
32. Verbeke W, Dejaeger K, Martens D, Hur J, Baesens B (2012) New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. *Eur J Oper Res* 218(1):211–229. <https://doi.org/10.1016/j.ejor.2011.09.031>
33. Verbraken T, Verbeke W, Baesens B (2012) A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE Trans Knowl Data Eng* 25(5):961–973. <https://doi.org/10.1109/TKDE.2012.50>
34. Óskarsdóttir M, Bravo C, Verbeke W, Sarraute C, Baesens B, Vanthienen J (2017) Social network analytics for churn prediction in telco: model building, evaluation and network architecture. *Expert Syst Appl* 85:204–220. <https://doi.org/10.1016/j.eswa.2017.05.028>
35. Zhu B, Baesens B, vanden Broucke SK (2017) An empirical comparison of techniques for the class imbalance problem in churn prediction. *Inf Sci* 408:84–99. <https://doi.org/10.1016/j.ins.2017.04.015>
36. Wael Fujo S, Subramanian S, Ahmad Khder M (2022) Customer churn prediction in telecommunication industry using deep learning. *Inf Sci Lett* 11(1):24. <https://doi.org/10.18576/isl/110120>
37. Alzubaidi AMN, Al-Shamery ES (2020) Projection pursuit Random Forest using discriminant feature analysis model for churners prediction in telecom industry. *Int J Electr Comput Eng*. <https://doi.org/10.11591/ijece.v10i2.pp1406-1421>
38. Amin A, Al-Obeidat F, Shah B, Adnan A, Loo J, Anwar S (2019) Customer churn prediction in telecommunication industry using data certainty. *J Bus Res* 94:290–301. <https://doi.org/10.1016/j.jbusres.2018.03.003>
39. Fathian M, Hoseinpoor Y, Minaei-Bidgoli B (2016) Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes*. <https://doi.org/10.1108/K-07-2015-0172>
40. Azeem M, Usman M, Fong ACM (2017) A churn prediction model for prepaid customers in telecom using fuzzy classifiers.

- Telecommun Syst 66(4):603–614. <https://doi.org/10.1007/s11235-017-0310-7>
41. Brandusoiu I, Todorean G (2013) Churn prediction in the telecommunications sector using support vector machines. *Margin* 1(1)
 42. Hung SY, Yen DC, Wang HY (2006) Applying data mining to telecom churn management. *Expert Syst Appl* 31(3):515–524
 43. Coussement K, Lessmann S, Verstraeten G (2017) A comparative analysis of data preparation algorithms for customer churn prediction: a case study in the telecommunication industry. *Decis Support Syst* 95:27–36. <https://doi.org/10.1016/j.dss.2016.11.007>
 44. Praseeda CK, Shivakumar BL (2021) Fuzzy particle swarm optimization (FPSO) based feature selection and hybrid kernel distance based possibilistic fuzzy local information C-means (HKD-PFLICM) clustering for churn prediction in telecom industry. *SN Appl Sci* 3:1–18. <https://doi.org/10.1007/s42452-021-04576-7>
 45. Denmark Statistics. (2023). Population figures. <https://www.dst.dk/en/Statistik/emner/borgere/befolkning/befolkningstal>. Accessed 19 Apr, 2023
 46. Kaggle Data Platform. (2017). Telco Customer Churn. <https://www.kaggle.com/datasets/blaschar/telco-customer-churn?sortBy=hotness&group=everyone&pageSize=20&datasetId=13996&language=Python>. Accessed 2 Oct, 2022
 47. Maven Analytics. (2022). DATA PLAYGROUND: Telecom Customer Churn. <https://www.mavenanalytics.io/data-playground>. Accessed 18 Oct, 2022
 48. Kaggle Data Platform. (2021). Cell2Cell Duke University Telco Dataset. <https://www.kaggle.com/datasets/geoamins/cell2cell-duke-university-telco-dataset>. Accessed 2 Oct, 2022
 49. Idris A, Rizwan M, Khan A (2012) Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Comput Electric Eng* 38(6):1808–1819. <https://doi.org/10.1016/j.compeleceng.2012.09.001>
 50. Japkowicz N, Stephen S (2002) The class imbalance problem: a systematic study. *Intel Data Anal* 6(5):429–449. <https://doi.org/10.3233/IDA-2002-6504>
 51. Jolly K (2018) Machine learning with scikit-learn quick start guide: classification, regression, and clustering techniques in Python. Packt Publishing Ltd. ISBN: 1789347378, 9781789347371
 52. Lalwani P, Mishra MK, Chadha JS, Sethi P (2022) Customer churn prediction system: a machine learning approach. *Computing* 104(2):271–294. <https://doi.org/10.1007/s00607-021-00908-y>
 53. Wade C (2020) Hands-on gradient boosting with XGBoost and scikit-learn: perform accessible machine learning and extreme gradient boosting with Python. Packt Publishing Ltd. ISBN: 1839213809, 9781839213809
 54. Svendsen GB, Prebensen NK (2013) The effect of brand on churn in the telecommunications sector. *Eur J Mark* 47(8):1177–1189. <https://doi.org/10.1108/03090561311324273>
 55. Alalwan AA, Baabdullah AM, Rana NP, Tamilmani K, Dwivedi YK (2018) Examining adoption of mobile internet in Saudi Arabia: extending TAM with perceived enjoyment, innovativeness and trust. *Technol Soc* 55:100–110. <https://doi.org/10.1016/j.techsoc.2018.06.007>
 56. Turel O, Serenko A (2006) Satisfaction with mobile services in Canada: an empirical investigation. *Telecommun Policy* 30(5–6):314–331. <https://doi.org/10.1016/j.telpol.2005.10.003>
 57. Bahri-Ammari N, Bilgihan A (2017) The effects of distributive, procedural, and interactional justice on customer retention: an empirical investigation in the mobile telecom industry in Tunisia. *J Retail Consum Serv* 37:89–100. <https://doi.org/10.1016/j.jretconser.2017.02.012>
 58. De Bock KW, De Caigny A (2021) Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling. *Decis Support Syst* 150:113523. <https://doi.org/10.1016/j.dss.2021.113523>
 59. Srinuan P, Srinuan C, Bohlin E (2014) An empirical analysis of multiple services and choices of consumer in the Swedish telecommunications market. *Telecommun Policy* 38(5–6):449–459. <https://doi.org/10.1016/j.telpol.2014.03.002>
 60. Ahmad R, Buttle F (2002) Customer retention management: a reflection of theory and practice. *Market Intell Plan* 20(3):149–161. <https://doi.org/10.1108/02634500210428003>
 61. Jyh-Fu Jeng D, Bailey T (2012) Assessing customer retention strategies in mobile telecommunications: hybrid MCDM approach. *Manag Decis* 50(9):1570–1595. <https://doi.org/10.1108/00251741211266697>
 62. Jørgensen S (2021) Vodafone-chef: Teleselskaber i Europa er nødt til at konsolidere. <https://24tech.dk/nyheder/telekommunikation/vodafone-chef-teleselskaber-i-europa-er-noedt-til-at-konsolidere/>. Accessed 23 Oct, 2022
 63. Breinstrup T (2015) Telia og Telenor dropper kæmpe fusion i Danmark. <https://www.berlingske.dk/virksomheder/telia-og-telenor-dropper-kaempe-fusion-i-danmark>. Accessed 18, 2022
 64. Odde UJ (2022). Telia går i offensiven: Ny brandstrategi skal sikre ny position. <https://markedsforing.dk/artikler/nyheder/telia-gaar-i-offensiven-ny-brandstrategi-skal-sikre-ny-position/>. Accessed 5 Nov, 2022
 65. Jensen D (2021) Telenor vil væk fra aggressiv fokus på priser - satser på nye forretningsområder. <https://www.computerworld.dk/art/257116/telenor-vil-vaek-fra-aggressiv-fokus-paa-priser-satser-paa-nye-forretningsomraader>. Accessed 7 Nov, 2022
 66. Olsen L (2022) Telenor leverer flot regnskab - men mister 29.000 mobilkunder. <https://24tech.dk/nyheder/telekommunikation/telenor-leverer-flot-regnskab-men-mister-29-000-mobilkunder/>. Accessed 17 Nov, 2022
 67. Telenor (2021) Telenor leverer kundevækst og den højeste indtjening i otte år. <https://www.mynewsdesk.com/dk/telenor/pressreleases/telenor-leverer-kundevaekst-og-den-hoejeste-ste-indtjening-i-otte-aar-3096283>. Accessed 7 Nov 2022
 68. Keramati A, Ardabili SM (2011) Churn analysis for an Iranian mobile operator. *Telecommun Policy* 35(4):344–356. <https://doi.org/10.1016/j.telpol.2011.02.009>
 69. Üstüner T, Holt DB (2010) Toward a theory of status consumption in less industrialized countries. *J Consum Res* 37(1):37–56. <https://doi.org/10.1086/649759>
 70. Erhvervsstyrelsen. (2011). Økonomiske nøgletal for telebranchen 2011. https://ens.dk/sites/ens.dk/files/Tele/okonomiske_noegletal_telebranchen_20111.pdf. Accessed 9 Nov, 2022
 71. Erhvervsstyrelsen. (2015). Økonomiske Nøgletal for Telebranchen 2015. https://ens.dk/sites/ens.dk/files/Tele/oekonomiske_noegletal_for_telebranchen_-_2015.pdf. Accessed 9 Nov, 2022
 72. Li H, Wu D, Li GX, Ke YH, Liu WJ, Zheng YH, Lin XL (2015) Enhancing telco service quality with big data enabled churn analysis: infrastructure, model, and deployment. *J Comput Sci Technol* 30(6):1201–1214. <https://doi.org/10.1007/s11390-015-1594-2>
 73. Statista (2022) Telecommunications in Denmark. <https://www-statista-com.zorac.aub.aau.dk/study/84839/telecommunications-industry-in-denmark/>. Accessed 23 Oct 2022

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