

Deep learning for customer churn prediction in e-commerce decision support

Maciej Pondel¹[\[https://orcid.org/0000-0002-1978-6571\]](https://orcid.org/0000-0002-1978-6571), Maciej Wuczyński²[\[https://orcid.org/0000-0001-7376-1933\]](https://orcid.org/0000-0001-7376-1933),
Wiesława Gryncewicz³[\[https://orcid.org/0000-0003-1208-4099\]](https://orcid.org/0000-0003-1208-4099), Łukasz Łysik⁴[\[https://orcid.org/0000-0002-7886-5130\]](https://orcid.org/0000-0002-7886-5130),
Marcin Hernes⁵[\[https://orcid.org/0000-0002-3832-8134\]](https://orcid.org/0000-0002-3832-8134), Artur Rot⁶[\[https://orcid.org/0000-0002-7281-8253\]](https://orcid.org/0000-0002-7281-8253), and Agata
Kozina⁷[\[https://orcid.org/0000-0003-0447-4038\]](https://orcid.org/0000-0003-0447-4038)

¹⁻⁷ Wrocław University of Economics and Business, Komandorska 118/120, 53345 Wrocław, Poland

Abstract. Churn prediction is a Big Data domain, one of the most demanding use cases of recent time. It is also one of the most critical indicators of a healthy and growing business, irrespective of the size or channel of sales. This paper aims to develop a deep learning model for customers' churn prediction in e-commerce, which is the main contribution of the article. The experiment was performed over real e-commerce data where 75% of buyers are one-off customers. The prediction based on this business specificity (many one-off customers and very few regular ones) is extremely challenging and, in a natural way, must be inaccurate to a certain extent. Looking from another perspective, correct prediction and subsequent actions resulting in a higher customer retention are very attractive for overall business performance. In such a case, predictions with 74% accuracy, 78% precision, and 68% recall are very promising. Also, the paper fills a research gap and contributes to the existing literature in the area of developing a customer churn prediction method for the retail sector by using deep learning tools based on customer churn and the full history of each customer's transactions.

Keywords: Churn prediction, deep learning, machine learning, e-commerce, decision support

Introduction

E-commerce has provided many new opportunities to consumers, and it is still opening new ones. The rapid expansion of IT technologies and the Internet has resulted in this sector's rapid growth. According to [1], the value of the e-commerce market in Poland exceeds 51 billion PLN, and 28 million Poles use online shopping in various forms. The global turnover of the e-commerce industry currently amounts to 3 trillion USD per year, and the fastest-growing market is the Asian market, whose dynamics reaches almost 30%. For comparison, the value of the European e-commerce channel was estimated at over 602 billion EUR, of which nearly 60% was generated by three markets: British, German, and French.

The success of companies hugely depends on how well they can analyze the data on their clients' behavior. Customers' churn may be considered as a lost opportunity for profit. The costs of gaining new customers are usually five to even six times higher than the costs of retaining an existing customer [2]. As a result, efforts made by marketing specialists to sustain market share have switched from focusing on acquiring new customers to retaining existing ones – reducing customer churn. For this reason, customer churn, also known as customer turnover, customer attrition, or customer deflection, is a major concern for a number of industries. This is particularly important in the e-commerce context, where

consumers are able to compare products or services and change the vendor with minimal effort.

Churn prediction is a Big Data domain, one of the most demanding use cases of recent time. It is also one of the most critical indicators of a healthy growing business, irrespective of the size or channel of sales. Customer attrition allows specialists to estimate the number of customers who will give up on the company's product or service subscription in a given time frame.

According to Ph. Kotler [3], companies annually lose 10 to 30 per cent of customers while acquiring new customers is about ten times costlier than maintaining existing ones. This information-rich sentence indicates how valuable customers are for a business. Research done by Amy Gallo [4] states that depending on the industry, acquiring a new customer is anywhere from 5-25% more expensive than retaining an existing one. So, it is essential to keep customers happy. For example, the telecommunications industry experiences an average 30-35 per cent annual churn rate; additionally, it costs 5-10 times more to gain a new customer than to retain an existing one [5].

It is crucial for a contemporary business to start analyzing why customers abandon relationships with a company by cancelling services or ceasing to buy products. This type of analysis allows e-commerce specialists to modify their current activities and adjust offers so that customer's needs are better covered, resulting in a lower churn rate.

This paper aims to develop a deep learning model for customers' churn prediction in e-commerce. The study pertains to the prediction of customer churn in B2C e-commerce. It also fills a research gap and contributes to the existing literature in the area of developing a customer churn prediction method for the retail sector by using deep learning tools based on customer churn and the full history of each customer's transactions, which is the major contribution of the article.

Related works

In the field of e-commerce, customer churn can be placed among the most critical problems that need to be addressed and thoroughly examined.

Compared to traditional shopping in retail stores, e-commerce has a significant advantage: instant and accurate track of records and in-depth data collection (shopping activities, order information, delivery information, billing address, etc.). This data collection allows multidimensional analysis of both customers and their buying habits, additionally helping businesses to treat customers as individuals and in a personalized manner. With the support of gathered data, it is possible to create customer-centric business intelligence based on the following business concerns [6]:

- Which subpages did the customer visit? How long did they stay there? What was the sequence in which they browsed a given web page?
- Who are the most/least valuable customers? What are their distinctive characteristics?
- Who are the most/least loyal customers, and how are they characterized?
- What are customers' purchase behavior patterns?
- Which types of customers are more likely to respond to a particular promotion?
- And so on

There are many both academia researchers and practitioners who have been actively trying to predict customer churn with the help of gathered data – statistics, data mining, or machine learning strategies.

Customer churn is a buzzword that has been used for a long time in the field of e-commerce, and early determination of consumers that might be lost should be identified

accurately through data mining and data analysis and related in time with effective marketing measures [7].

Churn models are made to detect, as soon as possible, signals of potential churn and help to identify customers willing to abandon a given company voluntarily.

Customer churn prediction is a very demanding and challenging process aimed at identifying consumers willing to abandon a company or a service. Decision-makers and machine learning specialists focus on designing models which can help to identify early churn signals and recognize consumers on the verge of a decision between leaving or continuing. Therefore, to retain customers, academics, as well as practitioners, find it crucial to build a churn prediction model that is as accurate as possible in order to minimize the risk of customer churn [8]. Also, researchers have confirmed that customer churn prediction models can improve a company's revenue and its reputation in the market [9], [10]. Reducing the rate of churn and retaining current customers are the most cost-effective marketing approaches that will maximize the shareholder's value [11], [12]. Today, companies have enough information, of every kind, about the behaviour of their customers - this has created an opportunity for the machine learning (ML) community to develop predictive modelling techniques to handle the customer churn prediction. [13].

During the last decade, customer churn prediction has received a growing consideration in order to survive in an increasingly competitive and global marketplace [14]. Companies should strive for models that can accurately identify potential churners, and this becomes even more important in the digital economy context. Over the last decade, this issue has been mentioned and researched by many practitioners and academics. In contemporary literature, we can observe two main trends concerning customer churn. According to [15], the first branch includes traditional classification methods such as decision tree (DT) and logistic regression (LR) [16] [17] [18] [19] [20] [21] [22] [23]. The second mentioned line of thought is based on artificial intelligence methods such as neural networks [24] [17] [23]; [25], evolutionary learning [17], genetic algorithms [17][18], random forests [26]; [27], improved balanced random forests [28], K-nearest neighbour [29], fuzzy logic Systems [30], and support vector machines [28]. The decision tree and logistic regression are dedicated to the analysis of continuous data; they cannot, however, guarantee the accuracy of constructed models for large scale, nonlinearity, and high-dimensionality [31].

All of the presented models of customer churn prediction are very helpful in creating measures which can help a company to prevent customers from attrition. Worth mentioning is that customer churn predictive models are usually solely evaluated based on their predictive performance in which the models show the ability to correctly identify customer churns and non-churns separately and accurately [21].

For the customer churn prediction problem, most of the related academic works focus on the so-called post-paid industries. This means that the contract with the customer ends or is terminated (e.g. banks, Internet service providers, insurance companies, and telephone service companies) [32] [13] [33]. The subject of this paper, as it was mentioned in the introduction, is the use of deep learning algorithms for churn prediction in the retail industry. A characteristic feature of this sector is the uncertainty surrounding the return of a customer to the same seller. As they are not bound by any contract, they can easily abandon the existing relationship. That is why the purpose of this article is to create a model that calculates the probability that a customer will return to the same vendor and in how many days they will return.

It becomes very significant in the e-commerce context, where competitors are only a few 'mouse clicks' away, and consumers can compare and contrast competing products and services with minimal expenditure of personal time or effort and move from one company to another [34]. In our deliberations, we will narrow down the research area even further, focusing on only one part of e-commerce, namely the retail sector. We are going to develop a useful churn prediction model for B2C context outperforming the commonly used methods because of two reasons. It is a model capable of capturing the specific characteristics of B2C

e-commerce relations, and – the second thing – it can predict when the customer will return to the same vendor.

In most domains, churning is usually referred to as losing a client. For example, [35] predicts the churn probability for prepaid clients of a cellular telecommunication company. In financial services (banking and insurance), churn is usually seen as closing accounts [32]. [36] predict the switching probability of an insured person to another auto insurance company. As far as retail is concerned, most studies also focus on the customer's ability to leave to identify the exact moment when customers will discontinue their relationship with companies. In the retail sector literature, churn has also been considered as the partial or progressive defection of customers. [37] used several classification techniques and proposed predictive models for partial customer turnover in retail. Most customers exhibit partial defection, which may subsequently lead to a complete switch. Also, Buckinx and Van den Poel [26] used the concept of partial churn to identify customers that the company should focus on if concerned with customer retention. The costs of gaining new customers are usually five to even six times higher than the costs of retaining an existing customer [2]. Nevertheless, they still talked about attrition. Also churn models based on risks models has been developed [38].

Our study pertains to the prediction of customer churn in B2C e-commerce. In contrast to existing research, we developed a deep learning model based on the full history of each customer's transactions, which can be useful in existing customer segmentation mechanism.

Materials and methods

Dataset and data processing

The original dataset consists of 626,275 rows and 131 columns. Each row concerns a single purchase and an aggregated history of all previous purchases of the customer who made it. The target variable, churn, indicates whether another purchase will be made by the same customer in the future. Preprocessing, conducted using Pandas 0.25.1 library installed under Python 3.7.4, included the removal of duplicates, redundant columns, and outliers. Principal component analysis was used as a dimensionality reduction technique to represent highly correlated variables ($\text{abs}(\text{Spearman correlation}) > 0.8$). After those steps, class imbalance was very high, as the data consisted of 79% of rows with churn=1. Thus, random undersampling was used to achieve class balance. Finally, the data, having 152,456 rows and 113 columns, (112 predictors: *base_price, discount, n_products, previous_Winter_hats, previous_Football_accessories, previous_Dresses, previous_DKNY, previous_Polo_shirts, previous_Training_shoes, previous_Stripes, previous_Lifestyle_shoes, previous_Swimsuits, previous_Care_products, previous_Wallets, previous_Gloves_and_scarves, previous_Running_shoes, previous_Backpacks, previous_Gucci, previous_Hilfiger, previous_Winter_coats, previous_Bags, previous_Shirts, previous_Casual_shoes, previous_Skirts, previous_Tops, previous_Trainers, previous_Balls, previous_Vests, previous_Basketball_shoes, previous_Jeans, previous_Sandals, previous_Underwear, previous_Tennis_shoes, previous_Outdoor_shoes, previous_Autumn_jackets, previous_Accessories, previous_Tracksuits, previous_Slippers, previous_Hiking_boots, previous_Trousers, previous_Glasses, previous_Training_accessories, previous_Sweaters, previous_Sweatshirts, previous_Shoes, previous_Shorts, previous_Clothes, previous_Hats, previous_Football_boots, previous_Armani, previous_Football_clothing, previous_Fleece, previous_Versace, previous_Socks, previous_Calvin_Klein, previous_Lauren, Shoes, Lauren, Running_shoes, Calvin_Klein, DKNY, Hilfiger, Clothes, Hats, Gucci, Lifestyle_shoes, Accessories, Tops, Sweatshirts, Trousers, Shorts, Versace, Casual_shoes, Armani, Stripes, Shirts, Skirts, Tennis_shoes, Glasses, Backpacks, Trainers, Socks, Slippers, Winter_hats, Autumn_jackets, Sandals, Bags, Football_accessories, Balls, Training_shoes, Outdoor_shoes, Football_boots, Basketball_shoes, Hiking_boots, Care_products, Underwear, Wallets, Winter_coats, Fleece, Tracksuits, Gloves_and_scarves, Swimsuits, Polo_shirts, Football_clothing, Sweaters, Dresses, Jeans, Training_accessories, Vests, transaction_date, days_since_fir*

st_purchase, value_previous_transaction, and target variable: churn), was standardized to be centered at zero and to have a unit variance.

In order to use a recurrent network topology, the data needed to be represented as a time series. Therefore, each row was transformed into a two-step series, with the first step including data on previous purchases and the second – on the current purchase. Some variables, such as the date of the purchase, were inadequate for such a transformation and were represented as two steps with identical values. The target variable was stored as a separate, 1-dimensional vector. The new dataset consisted of 152,456 time series with 2 steps and 58 features.

Model tuning

The prediction of customer churn was performed using two base artificial neural network topologies. A multilayer perceptron (MLP), with one or two fullyconnected dense layers was used. Also recurrent layer as a first hidden layer (RNN), optionally supported by an additional dense layer was used. Particular numbers of neurons were preliminarily selected by comparing the accuracy and F1 scores of models of different widths. When multiple models performed similarly, the simpler one was selected. Each network was optimized using a binary cross-entropy loss function. The output layer used a sigmoid activation function. For the purpose of overfitting prevention, each model was augmented with an extra dropout after every hidden layer. Both versions of the model, with and without dropout, were trained and compared. In case the “dying ReLU” problem appeared, each model was trained in two versions – using the standard rectified linear unit activation function and using the Leaky ReLU activation function. All models were prepared and trained using Keras 2.3.1 library with TensorFlow 2.0.0 backend [www.tensorflow.org].

Learning

A 10-fold split was performed over a dataset. Each model was trained independently 10 times, using consecutive data sections as validation sets and the remaining parts as training sets. The batch size amounted to 10,000 randomly selected rows. The models consisting of only fully-connected layers were trained over 40 epochs. The models containing recurrent layers were trained over 60 epochs. Model accuracy on the training and validation set was measured after each epoch. After the last epoch, additional metrics were calculated.

Experimental results

The trained models were used to predict samples from the validation set. Based on these predictions, a set of metrics was calculated and presented in Table 1. Accuracy, precision, recall, and confusion matrices were calculated for the prediction threshold equaling 0.5. For each row, the metrics were averaged over 10 independent models of the same architecture (but different dataset split using the 10 folds). The first column describes the number, type, and width of hidden layers. The second column indicates the probability of dropout for each hidden layer. The third column concerns the activation function used. The next columns contain the averaged metric values and their standard errors, in parentheses. The last column presents an averaged confusion matrix of the model, denoted using estimated probability values in percentages.

Table 1. Topologies of the models and corresponding metric values.

Network architecture	Dropout	Activation function	Average metric value ± SE					Averaged confusion matrix (%)
			Train accuracy	Test accuracy	Precision	Recall	AROC	
Dense(4)	0.0	ReLU	0.737 (0.001)	0.736 (0.001)	0.769 (0.003)	0.674 (0.004)	0.803 (0.002)	40 10 16 34
Dense(4)	0.3	ReLU	0.734 (<0.001)	0.733 (0.001)	0.742 (0.001)	0.714 (0.004)	0.807 (0.001)	38 12 14 36
Dense(4)	0.0	Leaky ReLU	0.735 (0.001)	0.735 (0.001)	0.765 (0.003)	0.678 (0.005)	0.807 (0.001)	40 10 16 34
Dense(4)	0.3	Leaky ReLU	0.732 (<0.001)	0.732 (0.001)	0.775 (0.002)	0.655 (0.003)	0.806 (0.001)	40 10 17 33
Dense(4), Dense (2)	0.0	ReLU	0.736 (0.001)	0.735 (0.001)	0.771 (0.002)	0.669 (0.002)	0.807 (0.002)	40 10 17 33
Dense(4), Dense (2)	0.3	ReLU	0.731 (<0.001)	0.730 (0.001)	0.733 (0.006)	0.726 (0.008)	0.805 (0.001)	37 13 14 36
Dense(4), Dense (2)	0.0	Leaky ReLU	0.737 (0.001)	0.736 (0.001)	0.770 (0.002)	0.672 (0.002)	0.810 (0.001)	40 10 16 33
Dense(4), Dense (2)	0.3	Leaky ReLU	0.733 (<0.001)	0.733 (0.001)	0.752 (0.002)	0.694 (0.002)	0.807 (0.001)	39 11 15 35
Recurrent(6)	0.0	ReLU	0.740 (0.001)	0.739 (0.001)	0.779 (0.002)	0.668 (0.002)	0.813 (0.001)	40 10 17 33
Recurrent(6)	0.3	ReLU	0.736 (0.001)	0.735 (0.001)	0.797 (0.003)	0.631 (0.002)	0.810 (0.002)	42 08 18 32
Recurrent(6)	0.0	Leaky ReLU	0.739 (0.001)	0.739 (0.002)	0.767 (0.003)	0.685 (0.002)	0.811 (0.002)	40 10 16 34
Recurrent(6)	0.3	Leaky ReLU	0.733 (<0.001)	0.732 (0.001)	0.780 (0.001)	0.647 (0.003)	0.807 (0.001)	41 09 18 32
Recurrent(6), Dense(4)	0.0	ReLU	0.739 (0.001)	0.739 (0.001)	0.778 (0.004)	0.668 (0.006)	0.812 (0.001)	40 10 17 33
Recurrent(6), Dense(4)	0.3	ReLU	0.739 (0.001)	0.737 (0.001)	0.777 (0.003)	0.665 (0.004)	0.811 (0.001)	40 10 17 33
Recurrent(6), Dense(4)	0.0	Leaky ReLU	0.739 (0.001)	0.739 (0.001)	0.777 (0.003)	0.669 (0.004)	0.812 (0.001)	40 10 17 33
Recurrent(6), Dense(4)	0.3	Leaky ReLU	0.733 (<0.001)	0.733 (0.001)	0.776 (0.002)	0.655 (0.002)	0.807 (0.001)	40 10 17 33

In order to compare the results of the 16 models, a series of statistical tests was performed, all with significance level $\alpha=0.05$. The introductory testing for normality, performed using Shapiro-Wilk test, revealed non-normality in some groups for every metric. A Levene test suggested a heterogeneity of variance of the training accuracy between the results of the models. The results of those tests implied further use of non-parametric methods. A Kruskal-Wallis test was used to check the equality of medians between the groups, while A Dunn's test with a Holm adjustment was applied for the post-hoc analysis. A Levene test with a Bonferonni correction was used for pairwise comparisons of variance. Simultaneous inference of multiple metrics was performed using a Friedman's test, with a Nemenyi test used for the post-hoc analysis.

The resulting models were of similar quality, with the global average accuracy of 73.6%. Accuracy was significantly less volatile than precision (Nemenyi $p=1.4 \times 10^3$) and recall ($p=2.6 \times 10^5$). AROC had lesser variance than recall, with marginal significance ($p=0.02$).

The Friedman test revealed a marginally significant difference of the combined metrics between the models ($p=0.01$). The Nemenyi post-hoc test suggested only one relevant difference ($p=0.05$), as the one-layer recurrent model without dropout and with ReLU activation performed better than one-layer MLP with 30% dropout and Leaky ReLU, considering all the measured metrics.

Although only two models were compellingly different in overall performance, there were 92 significant differences between particular metrics. The Dunn's test suggested 30 differences in the training accuracy between models. They overlapped with the only 5 significant differences in test accuracy. There were 27 relevant differences between models in precision and 26 – in recall. AROC values consistently differed between 4 models. Table 2 presents the results of the Dunn's test. Each pair of values indicate the number of models that performed, respectively, worse and better than the competitive model.

Table 2. Topologies of the models and aggregated results of the Dunn's post-hoc analysis.

Network architecture	Dropout	Activation function	Average metric value ± SE				
			Train accuracy	Test accuracy	Precision	Recall	AROC
Dense(4)	0.0	ReLU	+2, -0	+0, -0	+0, -1	+1, -0	+0, -2
Dense(4)	0.3	ReLU	+0, -3	+0, -0	+0, -8	+6, -0	+0, -0
Dense(4)	0.0	Leaky ReLU	+0, -0	+0, -0	+0, -1	+2, -0	+0, -0
Dense(4)	0.3	Leaky ReLU	+0, -7	+0, -0	+2, -0	+0, -4	+0, -0
Dense(4), Dense (2)	0.0	ReLU	+1, -0	+0, -0	+0, -1	+1, -0	+0, -0
Dense(4), Dense (2)	0.3	ReLU	+0, -8	+0, -5	+0, -8	+6, -0	+0, -2
Dense(4), Dense (2)	0.0	Leaky ReLU	+2, -0	+0, -0	+0, -1	+1, -0	+0, -0
Dense(4), Dense (2)	0.3	Leaky ReLU	+0, -5	+0, -0	+0, -6	+4, -0	+0, -0
Recurrent(6)	0.0	ReLU	+6, -0	+1, -0	+3, -0	+0, -0	+2, -0
Recurrent(6)	0.3	ReLU	+0, -0	+0, -0	+8, -0	+0, -9	+0, -0
Recurrent(6)	0.0	Leaky ReLU	+4, -0	+1, -0	+0, -1	+4, -0	+0, -0
Recurrent(6)	0.3	Leaky ReLU	+0, -5	+0, -0	+3, -0	+0, -5	+0, -0
Recurrent(6), Dense(4)	0.0	ReLU	+6, -0	+1, -0	+3, -0	+1, -2	+2, -0
Recurrent(6), Dense(4)	0.3	ReLU	+4, -0	+1, -0	+3, -0	+0, -2	+0, -0
Recurrent(6), Dense(4)	0.0	Leaky ReLU	+5, -0	+1, -0	+3, -0	+1, -0	+0, -0
Recurrent(6), Dense(4)	0.3	Leaky ReLU	+0, -2	+0, -0	+2, -0	+0, -4	+0, -0

Most models had a similar variance of all measured metrics. Precision was significantly less volatile ($p=0.03$) in the model with one recurrent and one dense layer, with 30% dropout and Leaky ReLU activations, than in the model with one dense layer, without dropout and with ReLU activation function. The model with two dense layers, 30% dropout, and Leaky ReLU has a lower variance of training accuracy than the same model without dropout ($p=0.03$), and than the recurrent model without a dense layer and dropout, with Leaky ReLU activation ($p=0.04$). Also, the model with one dense layer, 30% dropout, and Leaky ReLU activation had a significantly lower variance than the model with two dense layers, no dropout, and Leaky ReLU, with $p=0.04$.

Conclusions

The experiment was performed over real e-commerce data in an industry where 75% of buyers are one-off customers. It means that such a number of customers made a purchase only once and they have never returned to the store. In contrast, the number of regular customers (with more than 5 purchases) accounts for only 2% in the whole population. Such conditions in the e-commerce business make the input dataset unbalanced, which was mentioned in the specification of the method. It makes the churn prediction much more challenging than in any other line of business. The prediction basing on this business specificity (many one-off customers and very little regular ones), churn prediction is extremely challenging and in a natural way must be inaccurate to a certain level. Looking from another perspective, correct prediction and subsequent actions resulting with a higher customer retention are very attractive for the overall business performance. In such a case, prediction with 74% accuracy, 78% precision, and 68% recall is very promising. Even though in other business cases similar results could be considered insufficient, the achieved results

are significantly promising. The presented research has a preliminary status. The main disadvantage is using only a filter method of feature selection. The application of wrapper methods is needed for the reduction of the input attributes set. Also instead of using random sampling to generate the training and test datasets it might be interesting to develop approach to ensure that all transactions of a customer can exist only in one dataset. A very important issue also consists in the identification of the point in time at which the customer will return to the same retailer. Such an approach can better address the churn problem in a retail business due to the unclear definition of the churned customer. The research in these areas will be performed in the future works.

Funding: This research was funded by the Ministry of Science and Higher Education in Poland under the program "Regional Initiative of Excellence" 2019–2022, project number 015/RID/2018/19, total funding amount 10,721,040.00 PLN.

References

- [1] E-COMERCE Homepage.
<https://media.pl.cushmanwakefield.com.pl/pr/444970/deweloperzy-magazynowi-i-operatorzy-logistyczni-sa-zgodni-e-commerce-r>. Accessed 2021 January 04.
- [2] Bhattacharya CB. When Customers Are Members: Customer Retention in Paid Membership Contexts. *Journal of the Academy of Marketing Science*. 1998 01 01;26(1):31-44. <https://doi.org/10.1177/0092070398261004>
- [3] Kotler P, Keller KL. *Marketing Management*. 14th Edition. Pearson; 2012.
- [4] Gallo A. The Value of Keeping the Right Customers. *Harvard Business Review* (<https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>). 2014;
- [5] Lu J. Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS . SUGI 27., Paper 114-27.
- [6] Chen D, Sain SL, Guo K. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*. 2012 08 27;19(3):197-208.
<https://doi.org/10.1057/dbm.2012.17>
- [7] Li X, Li Z. A Hybrid Prediction Model for E-Commerce Customer Churn Based on Logistic Regression and Extreme Gradient Boosting Algorithm. *Ingénierie des systèmes d'information*. 2019 Nov 26;24(5):525-530. <https://doi.org/10.18280/isi.240510>
- [8] Gordini N, Veglio V. Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*. 2017 04;62:100-107.
<https://doi.org/10.1016/j.indmarman.2016.08.003>
- [9] Oskarsdottir M, Bravo C, Verbeke W, Sarraute C, Baesens B, Vanthienen J. A comparative study of social network classifiers for predicting churn in the telecommunication industry. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016 08.
<https://doi.org/10.1109/asonam.2016.7752384>
- [10] Amin A, Anwar S, Adnan A, Khan MA, Iqbal Z. Classification of cyber attacks based on rough set theory. *2015 First International Conference on Anti-Cybercrime (ICACC)*. 2015 First International Conference on Anti-Cybercrime (ICACC). 2015 Nov.
<https://doi.org/10.1109/anti-cybercrime.2015.7351952>
- [11] Ekinci Y, Uray N, Ülengin F. A customer lifetime value model for the banking industry: a guide to marketing actions. *European Journal of Marketing*. 2014 04 08;48(3/4):761-784.
<https://doi.org/10.1108/ejm-12-2011-0714>

- [12] Ngai E, Xiu L, Chau D. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*. 2009 03;36(2):2592-2602. <https://doi.org/10.1016/j.eswa.2008.02.021>
- [13] Amin A, Al-Obeidat F, Shah B, Adnan A, Loo J, Anwar S. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*. 2019 01;94:290-301. <https://doi.org/10.1016/j.jbusres.2018.03.003>
- [14] Gordini N. Market-Driven Management: A Critical Literature Review. *Symphonya. Emerging Issues in Management*. 2010 Dec 01;(2). <https://doi.org/10.4468/2010.2.08gordini>
- [15] Çelik Ö, Usame OO. Comparing to techniques used in customer churn analysis. *Journal of Multidisciplinary Developments*. 2019;4(1).
- [16] Burez J, Van den Poel D. CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*. 2007 02;32(2):277-288. <https://doi.org/10.1016/j.eswa.2005.11.037>
- [17] Gordini N, Veglio V. Using neural networks for customer churn prediction modeling: preliminary findings from the Italian electricity industry. *Proceedings de X° Convegno Annuale della Società Italiana Marketing: "Smart Life. Dall'Innovazione Tecnologica al Mercato", Università degli Studi di Milano-Bicocca, Italy*. 2013, 1–13.
- [18] Gordini N, Veglio V. Customer relationship management and data mining: A classification decision tree to predict customer purchasing behavior in global market. In: *Handbook of Research on Novel Soft Computing Intelligent Algorithms: Theory and Practical Applications*. Vol. 1-2. 2013:1–40.
- [19] Verbeke W, Martens D, Mues C, Baesens B. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*. 2011 03;38(3):2354-2364. <https://doi.org/10.1016/j.eswa.2010.08.023>
- [20] Verbeke W, Dejaeger K, Martens D, Hur J, Baesens B. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*. 2012 04;218(1):211-229. <https://doi.org/10.1016/j.ejor.2011.09.031>
- [21] De Caigny A, Coussement K, De Bock KW. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*. 2018 09;269(2):760-772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- [22] Deng Z, Lu Y, Wei KK, Zhang J. Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in China. *International Journal of Information Management*. 2010 08;30(4):289-300. <https://doi.org/10.1016/j.ijinfomgt.2009.10.001>
- [23] Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH. Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research*. 2006 05;43(2):204-211. <https://doi.org/10.1509/jmkr.43.2.204>
- [24] Xu S, Lai S, Qiu M. Privacy preserving churn prediction. *Proceedings of the 2009 ACM symposium on Applied Computing - SAC '09*. the 2009 ACM symposium. 2009. <https://doi.org/10.1145/1529282.1529643>
- [25] Sharma A, Kumar Panigrahi P. A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *International Journal of Computer Applications*. 2011 08 31;27(11):26-31. <https://doi.org/10.5120/3344-4605>
- [26] Buckinx W, Van den Poel D. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*. 2005 07;164(1):252-268. <https://doi.org/10.1016/j.ejor.2003.12.010>
- [27] Coussement K, Van den Poel D. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques.

- Expert Systems with Applications*. 2008 01;34(1):313-327.
<https://doi.org/10.1016/j.eswa.2006.09.038>
- [28] Xie Y, Li X, Ngai E, Ying W. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*. 2009 04;36(3):5445-5449.
<https://doi.org/10.1016/j.eswa.2008.06.121>
- [29] Ahmed AA, Maheswari D. Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal*. 2017 Nov;18(3):215-220.
<https://doi.org/10.1016/j.eij.2017.02.002>
- [30] Abbasimehr H, Setak M, Tarokh MJ. A Neuro-Fuzzy Classifier for Customer Churn Prediction. *International Journal of Computer Applications*. 2011 Apr;19(8):35-41.
- [31] Yu X, Guo S, Guo J, Huang X. An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*. 2011 03;38(3):1425-1430. <https://doi.org/10.1016/j.eswa.2010.07.049>
- [32] Huang Y, Kechadi T. An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*. 2013 Oct;40(14):5635-5647.
<https://doi.org/10.1016/j.eswa.2013.04.020>
- [33] Pelka M, Rybicka A. Identification of factors that can cause mobile phone customer churn with application of symbolic interval-valued logistic regression and conjoint analysis. The 13th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. 2019, 187–195.
- [34] Tamaddoni Jahromi A, Stakhovych S, Ewing M. Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*. 2014 Oct;43(7):1258-1268.
<https://doi.org/10.1016/j.indmarman.2014.06.016>
- [35] Owczarczuk M. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*. 2010 06;37(6):4710-4712.
<https://doi.org/10.1016/j.eswa.2009.11.083>
- [36] Hur Y, Lim S. Customer Churning Prediction Using Support Vector Machines in Online Auto Insurance Service. In: Wang J, Liao XF, Yi Z, eds. *Advances in Neural Networks – ISNN 2005*. 3497. Berlin, Heidelberg: Springer; 2005.
https://doi.org/https://doi.org/10.1007/11427445_149
- [37] Miguéis V, Van den Poel D, Camanho A, Falcão e Cunha J. Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*. 2012 09;39(12):11250-11256. <https://doi.org/10.1016/j.eswa.2012.03.073>
- [38] Farquad M, Ravi V, Raju SB. Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing*. 2014 06;19:31-40.
<https://doi.org/10.1016/j.asoc.2014.01.031>
- [39] Slof D, Frasinca F, Matsiako V. A competing risks model based on latent Dirichlet Allocation for predicting churn reasons. *Decision Support Systems*. 2021 07;146:113541.
<https://doi.org/10.1016/j.dss.2021.113541>