

Zatonatska, Tetiana; Fareniuk, Yana; Shpyrko, Viktor

## Article

# Churn rate modeling for telecommunication operators using data science methods

Marketing i menedžment inovacij

**Provided in Cooperation with:**

ZBW OAS

*Reference:* Zatonatska, Tetiana/Fareniuk, Yana et. al. (2023). Churn rate modeling for telecommunication operators using data science methods. In: Marketing i menedžment inovacij 14 (2), S. 163 - 173.  
[https://mmi.sumdu.edu.ua/wp-content/uploads/2023/06/A690-2023\\_15\\_Zatonatska-et-al-3.pdf](https://mmi.sumdu.edu.ua/wp-content/uploads/2023/06/A690-2023_15_Zatonatska-et-al-3.pdf)  
doi:10.21272/mmi.2023.2-015.

This Version is available at:  
<http://hdl.handle.net/11159/631414>

## Kontakt/Contact

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics  
Düsternbrooker Weg 120  
24105 Kiel (Germany)  
E-Mail: [rights\[at\]zbw.eu](mailto:rights[at]zbw.eu)  
<https://www.zbw.eu/>

## Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte. Alle auf diesem Vorblatt angegebenen Informationen einschließlich der Rechteinformationen (z.B. Nennung einer Creative Commons Lizenz) wurden automatisch generiert und müssen durch Nutzer:innen vor einer Nachnutzung sorgfältig überprüft werden. Die Lizenzangaben stammen aus Publikationsmetadaten und können Fehler oder Ungenauigkeiten enthalten.

## Terms of use:

*This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence. All information provided on this publication cover sheet, including copyright details (e.g. indication of a Creative Commons license), was automatically generated and must be carefully reviewed by users prior to reuse. The license information is derived from publication metadata and may contain errors or inaccuracies.*



<https://savearchive.zbw.eu/terms-of-use>

## CHURN RATE MODELING FOR TELECOMMUNICATION OPERATORS USING DATA SCIENCE METHODS

**Tetiana Zatonatska,**  <https://orcid.org/0000-0001-9197-0560>

Dr.Sc., Professor, Taras Shevchenko National University of Kyiv, Ukraine

**Yana Farenjuk,**  <https://orcid.org/0000-0001-6837-5042>

Taras Shevchenko National University of Kyiv, Ukraine

**Viktor Shpyrko,**  <https://orcid.org/0000-0003-4955-2685>

Ph.D., Taras Shevchenko National University of Kyiv, Ukraine

**Correspondence author:** Tetiana Zatonatska, [tzatonat@ukr.net](mailto:tzatonat@ukr.net)**Type of manuscript:** research paper

**Abstract.** *The telecommunication company functioned in the market with extremely high competitiveness. Attracting new customers needs 5-10 times more expenses than maintaining an existing one. As a result, effective customer churn management and analysis of the reasons for customer churn are vital tasks for telecommunication operators. As a result, predicting subscriber churn by switching on the competitors becomes very important. Data Science and machine learning create enormous opportunities for solving this task to evaluate customer satisfaction with company services, determine factors that cause disappointment, and forecast which clients are at a greater risk of abandoning and changing services suppliers. A company that implements data analysis and modelling to develop customer churn prediction models has an opportunity to improve customer churn management and increase business results. The purposes of the research are the application of machine learning models for a telecommunications company, in particular, the construction of models for predicting the user churn rate and proving that Data Science models and machine learning are high-quality and effective tools for solving the tasks of forecasting the key marketing metrics of a telecommunications company. Based on the example of Telco, the article contains the results of the implementation of various models for classification, such as logistic regression, Random Forest, SVM, and XGBoost, using Python programming language. All models are characterised by high quality (the general accuracy is over 80%). So, the paper demonstrates the feasibility and possibility of implementing the model to classify customers in the future to anticipate subscriber churn (clients who may abandon the company's services) and minimise consumer outflow based on this. The main factors influencing customer churn are established, which is basic information for further forecasting client outflow. Customer outflow prediction models implementation will help to reduce customer churn and maintain their loyalty. The research results can be useful for optimising marketing activity of managing the outflow of consumers of companies on the telecommunication market by developing effective decisions based on data and improving the mathematical methodology of forecasting the outflow of consumers. Therefore, the study's main theoretical and practical achievements are to develop an efficient forecasting tool for enterprises to control outflow risks and to enrich the research on data analysis and Data Science methodology to identify essential factors that determine the propensity of customers to churn.*

**Keywords:** churn, client, consumer, Data Science, forecasting, machine learning, marketing, modelling, outflow.

**JEL Classification:** M0, O1

**Received:** 21 December 2022

**Accepted:** 30 May 2023

**Published:** 30 June 2023

**Funding:** There is no funding for this research.

**Publisher:** Academic Research and Publishing UG, Germany.

**Founder:** Sumy State University and Academic Research and Publishing UG (i.G.), Germany

**Cite as:** Zatonatska, T., Farenjuk, Y., & Shpyrko, V. (2023). Churn Rate Modeling for Telecommunication Operators Using Data Science Methods. *Marketing and Management of Innovations*, 2, 163–173. <https://doi.org/10.21272/mmi.2023.2-15>



Copyright: © 2023 by the author. This open-access article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Introduction.** Modern digitalisation trends extend to all areas of people's lives and the economy. In today's global economy, working with data is a vital component of business strategy for business development and expansion. The continued expansion of telecommunication technologies during COVID-19 has increased competition between businesses and changed consumer behaviour forever, putting much pressure on businesses and opening new opportunities for them to compete locally and globally. This trend has accelerated the digitisation of the economy due to the growth of the importance of e-commerce in the current stage and the movement of large numbers of online consumers. This trend will continue to grow, leading to the fact that in the future, at least 90% of purchases will be made online, and retaining consumers will be an increasingly difficult task. At the same time, consumer standards have also increased, increasing pressure on businesses to adopt new technologies.

Big Data Analytics (BDA) is a process of analysis the big data, which is characterised by large volumes and higher value, that gives a vision for making effective management decisions about business development (Ryfiak, 2020). BDA, in many ways, determines the activities of companies because it contributes to understanding the behaviour and interests of users, identifying user satisfaction or the risk of their outflow, and increasing business revenues (Riddle, 2020). With this in mind, Data Science and Machine Learning methods, models, and technologies must be increasingly applied to improve operations and business development of enterprises. Therefore, companies must assess their relationships with clients and evaluate success from the client's perspective and carry out a customer-centric strategy. The advancement of technologies such as database processing technologies contributes to innovations that increase the interactivity and personalisation of the consumer experience (Rosario et al., 2021). This influences competition growth and drives enterprises to pay more attention to the needs of consumers, turning them into powerful elements of business strategy and business processes. Management decisions and operations should be underpinned by knowledge of current and potential consumers and systematic analysis of customer relationships. Realising consumer research based on the implementation of modern ICT tools supports customer profiling and determining the multiple behavioural dimensions that impact customers' behaviour and intentions (Singh et al., 2020). Currently, the telecommunication industry moves to saturation as supply overtakes demand (Fang, 2021), and the key source of business growth can be achieved by switching between providers. The marketing research demonstrates that attracting a new client needs 5-10 times more expenses than saving a current one. So, effective management of client outflow and determining the reasons for client churn have become very important tasks for telecommunication companies (Thakkar et al., 2022). So, forecasting consumer outflow is one of the main in the marketing activity of telecommunication company, which increase the relevance of current research and the necessity of finding appropriate methods for its solving.

The purpose of the current study is the application of different machine learning models for a telecommunications company, in particular, the construction of models for predicting the user churn rate and proving that Data Science models and machine learning are high-quality and effective tools for solving the tasks of forecasting the key marketing metrics of a telecommunications company.

**Literature Review.** Many research papers by modern scientists prove the relevance of using big data to analyse various processes. There has been growing attention to the BDA in the last ten years. With the rapid increase in the use of Internet technologies, a huge amount of generated and stored data (Sekli et al., 2021). Many companies use the analysis of this big data to improve their business strategy and gain market advantages (Maroufkhani et al., 2019) by improving their processes, improving the work with the customer base, increasing loyalty to the company, and maintaining and increasing revenues (Riddle, 2020). The main goal of BDA is the improvement of the decision-making processes through exploring and analysing big data, such as social media data (text, photo, video), text messages, etc. (Ryfiak, 2020). Implementing BDA helps companies improve their operational activities and develop effective strategies to gain benefits and improve customer value (Arya et al., 2016). Technologies such as BDA are required to generate opportunities to innovate by extracting valuable insights. Sekli and Vega (2021) examine the factors influencing the adoption of BDA and assess its relationships with knowledge and performance management. In addition, the authors provide practical guidelines for decision-makers, which are responsible for defining a strategy for the execution of big data in institutions of high education.

Govorek (2021) substantiates the strategies of telecom operators, which improve their competitive positions in the market and make it possible to attract a wide customer base through the use of BDA. Radukych et al. (2019) examine the impact and consequences of digital transformation on telecommunication industries, which have signs of network externalities. Additionally, the article presents a methodological basis for assessing the information society. In a paper by Yusuf-Asaiah et al. (2017), the authors offer to evaluate the

recognised quality of experience to help telecom operators manage network performance more effectively and provide satisfaction with mobile Internet services.

Mobile network operators are facing a huge increase in data on their networks. Big data technologies present an up-to-date approach to working with accumulated data. A study by Simakovich et al. (2021) proposes solutions for big data that can accumulate and proceed huge amounts of data to obtain valuable insights. It creates an opportunity for mobile operators to improve the quality of their networks and offer their customers complete IoT solutions. The significance of the analysis is that the attitude of customers towards the brands and the products they purchase has increased. Data analysis is a main tool for identifying customer sentiment, specific product or service problems, and extending customer lifetime value (Parveen et al., 2021). A study by Dai (2017) highlights the significance of consumer feedback, which is a great source of information for telecommunication companies. This article uses a quantitative methodology to predict the number of dissatisfied users. The author focuses on selecting prognostic features and then the classification model. Dai (2017) used several data sets for the research related to complaints and survey data.

The triumph of e-commerce is mainly supported by consumers' engagement in the Internet environment. Hsu (2019) shows that using a marketing strategy when consumers consider the results of participation or observation of an event is effective because it becomes an incentive and induces purchase intention. A customer-centric marketing strategy focuses on consumers' feelings, ideas, emotions, and behaviours that are reflected during interactions in the online environment. However, Yen (2014) claims that consumer behaviour and purchase intentions change due to the expansion of the Internet and e-commerce platforms based on the Internet. For example, today's customers are concerned about information satisfaction, an organisation's reputation, and enhanced benefits. Consequently, companies are combining social media and e-commerce to maximise exposure and growth brand visibility and awareness.

Marketing can be specified as profitably determining and satisfying customers' needs. In an article by Varga and Gabor (2021), they state that marketing on social networks helps companies more easily promote products and services by focusing advertising on the more appropriate consumer segments. The major hypothesis of the study about the statement that social networks influence people's decisions was confirmed.

Social media allows for conveying brand information and offers, and e-commerce platforms provide a place where clients can see all products and relevant product information to make decisions. Company website data can be actively used to improve the organisation's business strategy. Unlike traditional offline marketing, the development of e-commerce makes it possible for clients to receive relevant information about enterprises, their proposals, and business activities from websites, as well as reviews and recommendations from consumers (Rosario et al., 2021). As a result, it increases competition because the consumer can easily switch from one service provider to another. Increasing satisfaction due to upgraded service, the proposition of more qualitative products, interactions, and the formation of good experiences is the basis for the success of companies operating in the competitive environment.

E-commerce involves implementing digital technology in business processes to increase online transactions and sales. However, the evolution of e-commerce occurs under the change in technological development (Yang et al., 2019). Thanks to the company's data analysis technologies through the application of Data Science and Machine Learning, businesses can monitor demographic characteristics and consumer purchasing activity to match the products and services proposed by the company to consumer requirements and needs and predict key business performance indicators. Machine learning and Data Science technologies allow real-time monitoring of the relationship between the efficiency of the business and marketing activity (including media activity) (Chornous et al., 2021; Farenjuk, 2022). In general, such studies are highly relevant for making predictions and developing optimal decisions for business management in conditions of uncertainty and high risk.

Implementing modern technologies included in Data Science, such as Data Engineering, Machine Learning, and Data Management, is now crucial to process data (Moorthi et al., 2020) effectively. Hurtado et al. (2019) mentioned that it is important to form an effective process of collecting and analysing all possible business functions and implement it in a complete Data Science solution to improve and control business performance because, at the current stage of market development, it is not enough for enterprises only to analyse datasets to form highly accurate predictions, develop the most cost-effective processes and minimise costs (Fedirko et al., 2019).

In general, there are a huge number of Data Science tools, and they should be implemented in business processes. Nevertheless, the technologies and methodology of Data Science and Machine Learning are not enchantments for business, as implementation processes directly affect their effectiveness. Experts claim that around 85% of projects must be implemented more effectively. The correct management process for



implementing and developing Data Science projects is key. It is important to abide by all the universal rules for efficient work with innovation and effectively using technology in the company so that this innovation is useful (Zhang et al., 2021).

**Methodology and research methods.** In the current research, the following methods of general scientific research were applied: empirical (familiarisation with the basics of Data Science methods), theoretical, observation, experiment, analysis and synthesis, induction and deduction, as well as mathematical methods, statistical information processing methods, and parametric models.

Machine learning classification models were used to predict e-commerce metrics, in particular churn rate for telecommunication companies (Telco was selected for an experiment). The following four classical machine learning methods were chosen for analysis and simulation: boosting (XGBoost), logistic regression, classification trees on an example of Random Forest, and support vector machines (SVM). These methods are the most popular and the most useful considering the task and specific of collected data for the experiment.

We define the stages of machine learning as direct learning; evaluation of the quality of models; choosing the best of them; optimisation of input parameters in the model to increase accuracy; assessment of the target parameter or classification; analysis of results. By training a model, we understand the transfer of data to derive a certain pattern of dependencies between them. The data that the model accepts for analysis should be divided into 2 or 3 groups: training, test, and validation samples. When building a computer model, the software is selected for calculating the constructed mathematical models and obtaining results per the set goal. This research used Python programming languages for data processing and modelling.

To assess the accuracy of classification methods, the confusion matrix will be used accuracy, which shows the percentage of correctly grouped objects from the database; precision (the proportion of objects that were determined by the classifier as positive (belonging to the specific group under consideration) and, indeed, is positive), recall (responsible for the ability of the algorithm to correctly find the group as a whole and is defined as the share of objects of the positive group, which were correctly classified by the algorithm, to all objects of this group). These characteristics allow us to choose the optimal model for further analysis.

**Results.** In order to investigate the practical application of Data Science in e-commerce, we will use several models based on open data of a telecommunications company that provides several services to its users. The data research was conducted for a Telco and was downloaded from IBM Sample Data Sets for customer retention programs. The study aims to predict churn behaviour to help the company develop a customer retention strategy. Each row of the collected database represents a customer, and each column contains a factor that characterises it:

- customers with churn within the last month;
- the subscribed services by each customer, such as phone connection, Internet, streaming TV and movies, multiple lines, online security, device protection, online backup, and technical support;
- information about the client's account - how long he has been a client, type of contract, method of payment, monthly payments, paperless account, and general costs;
- information about demographic characteristics of the clients – age, gender, presence of partners.

Customer outflow is one of the larger problems for the telecommunications market. Studies have proved that the average monthly churn rate metric among leading mobile and internet operators in the US market is from 1.9% to 2%. Nevertheless, before building models for outflow forecasting, it is required to examine the dataset's structure and quality and form hypotheses.

By examining the correlations between variables and churn metric, we can conclude that monthly contracts, lack of technical support, and lack of online security positively correlate with churn. On the other hand, contracts with two-year subscriptions have a negative correlation with churn. Online services such as streaming TV, online security, technical support, online backup, etc. are negatively associated with outflow (Figure 1).

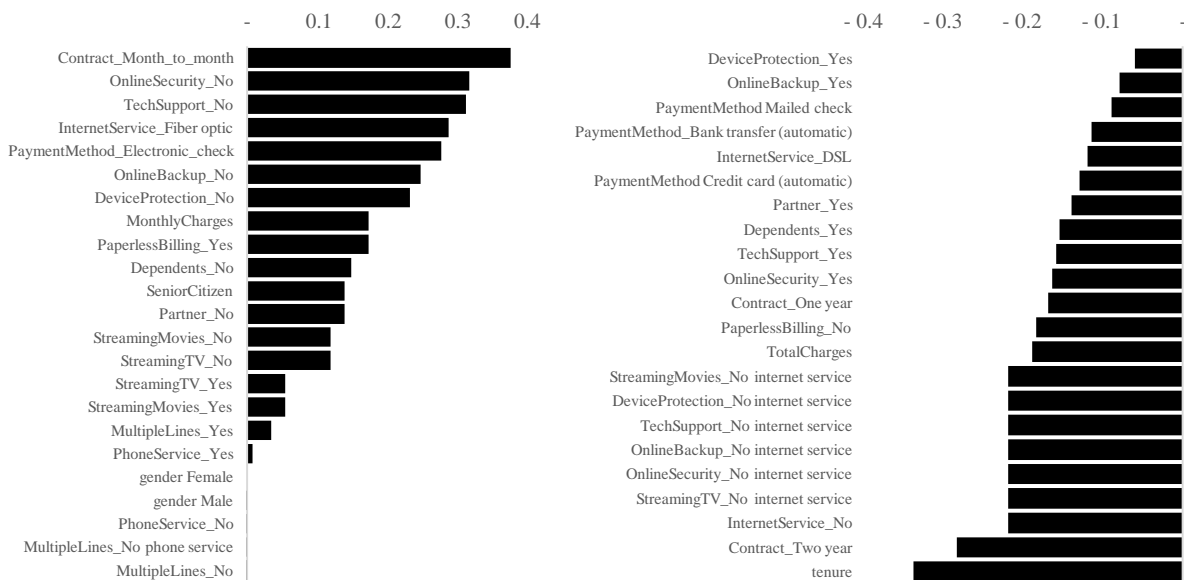
Many customers use the services of a telecommunications company for only a month, while quite a few use the company's services for around 72 months. This could be because different customers have various contracts. So, according to the contract they entered into, it may be more challenging for clients to leave a Telco or stay with the company.

The number of customers under different contracts should be considered. Most customers (almost 4,000) are on a monthly contract, while 1- and 2-year contracts have the same number of customers (1,500-2,000). Most monthly contracts are active during 1-2 months, while contracts for 2-years subscriptions tend to survive around 70 months. This displays that clients who take a longer contract have higher loyalty to the provider and tend to subscribe to services for a longer period.

Next, it is worth looking at the value of the main metric, which will need to be predicted. Churn rate - the outflow rate of consumers who refused to consume and use the company's services during a certain period. According to data, 26% of customers refuse to use the services of this telecommunications company. The data is distorted because a much larger part of customers does not stop using the services. It is critical to keep in mind in the process of modelling that inequality can lead to a lot of false negatives.

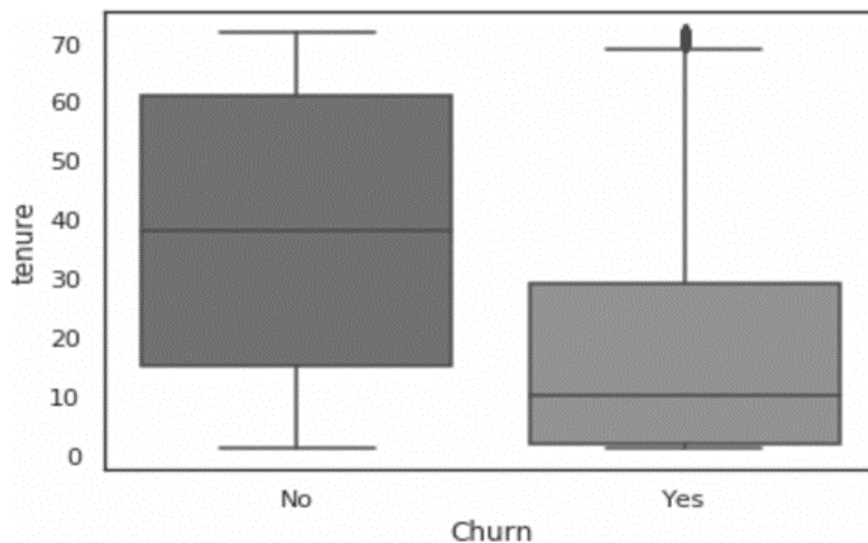
Let us consider the dependence of the outflow coefficient (churn rate) on some other variables and form hypotheses:

- Churning vs Tenure: As can be seen from Figure 2, non-churning customers tend to stay longer with a Telco.
- Churn by type of contracts: Customers with a monthly contract are characterised by a higher churn rate (Figure 3).
- Outflow by age: The attrition rate of the elderly is almost twice as high as that of the younger population (Figure 4).
- Churn by monthly and total payments: a higher level of customer churn when monthly or total payments are high (Figure 5 and Figure 6).



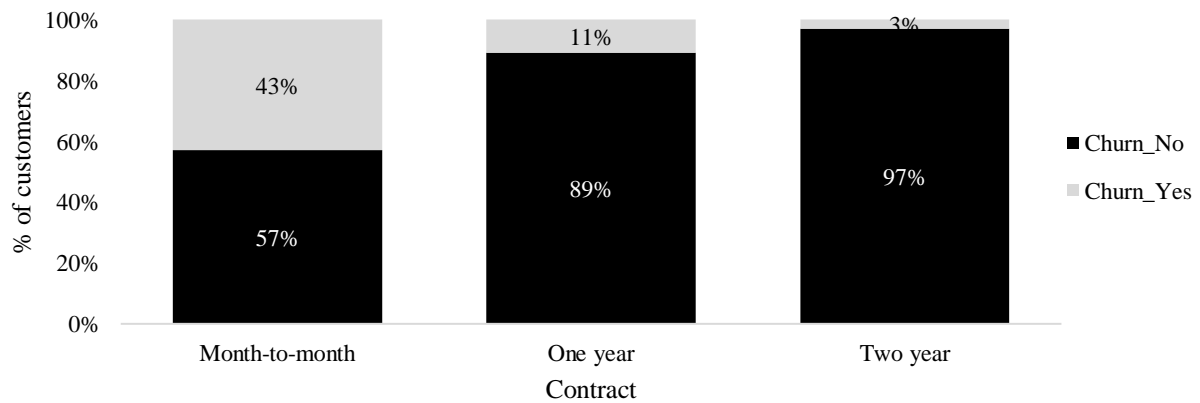
**Figure 1. Correlation of variables with a churn rate**

Sources: developed by the authors.



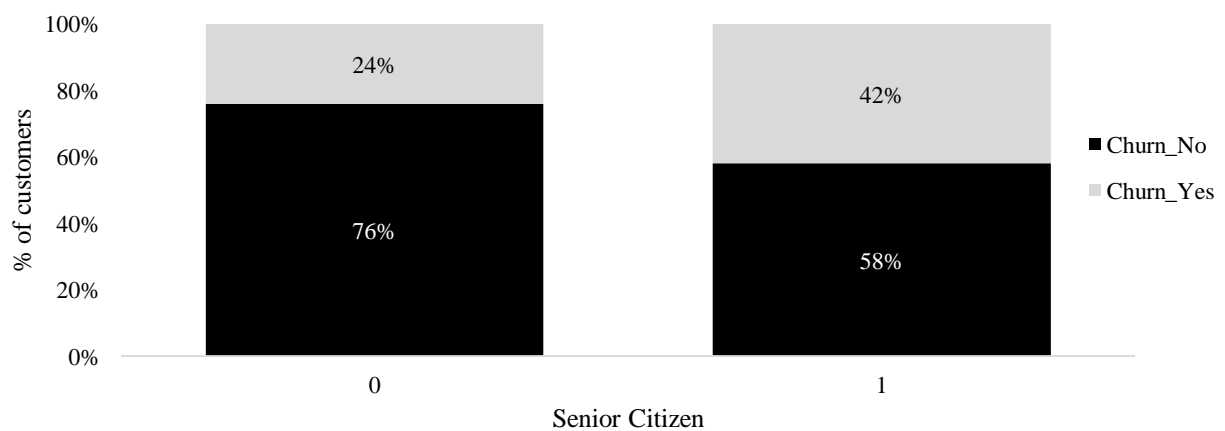
**Figure 2. Dependence of churn and tenure**

Sources: developed by the authors.



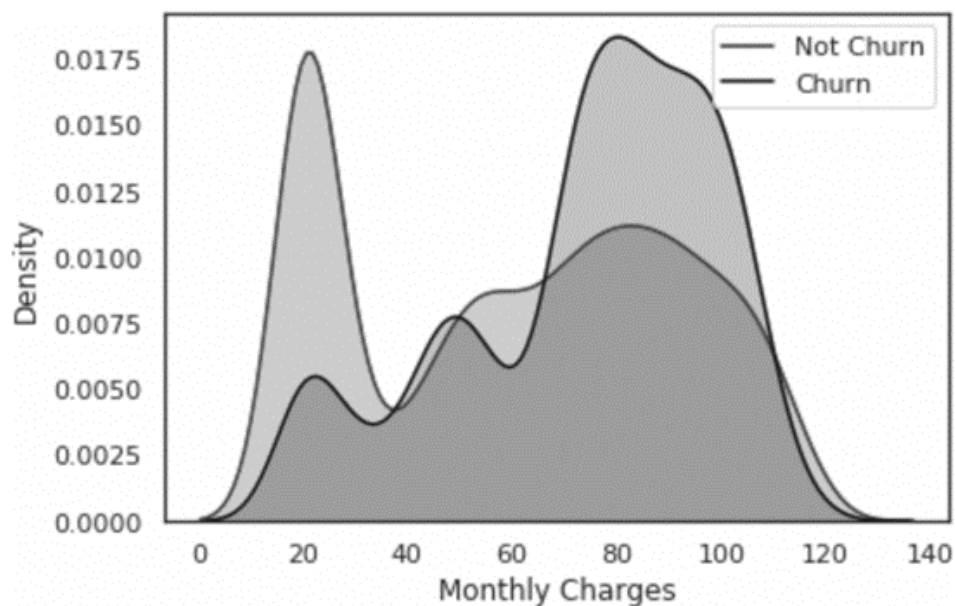
**Figure 3. Churn by contract type**

Sources: developed by the authors.



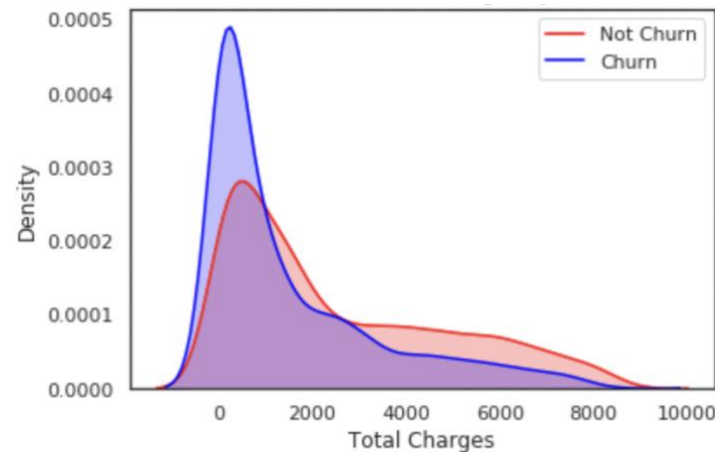
**Figure 4. Churn by seniority level (age of client was grouped in 2)**

Sources: developed by the authors.



**Figure 5. Distribution of monthly charges by churn**

Sources: developed by the authors.



**Figure 6. Distribution of total charges by churn**

Sources: developed by the authors.

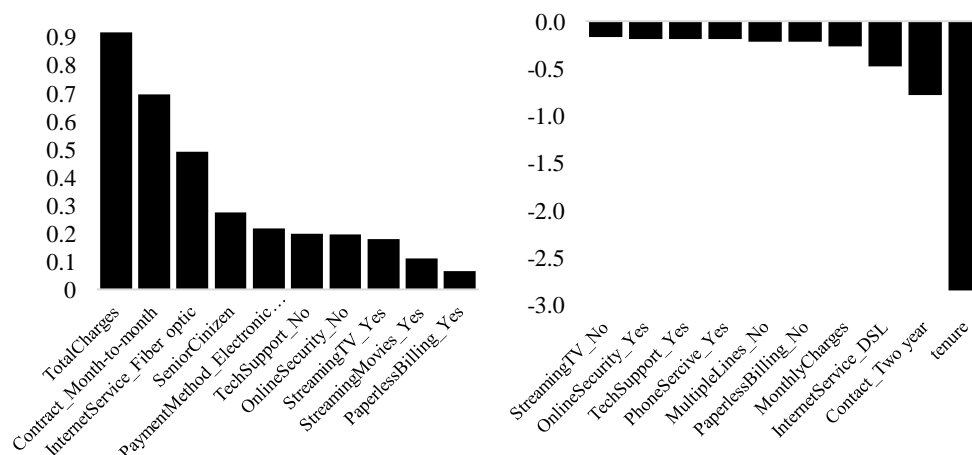
The study applies four models: Random Forest, XGBoost, logistic regression, and SVM. Logistic regression – it is crucial to scale all indicators in the logistic regression and achieve that they are all between 0 and 1. This helped increase the accuracy from 79.7% to 80.8%. Logistic regression has low false negative results (Table 1).

**Table 1. Confusion matrix gained from the implementation of logistic regression**

True label	Predicted label	
	0	1
0	1 418	162
1	244	286

Sources: developed by the authors.

By examining the effects of the factors (Figure 7), it can be seen that some characteristics have a positive relationship with the churn as a predicted variable, while others have a negative relationship. A negative correlation indicates that the probability of outflow drops with those characteristics.



**Figure 7. Influence of factors on customer churn, determined by logistic regression**

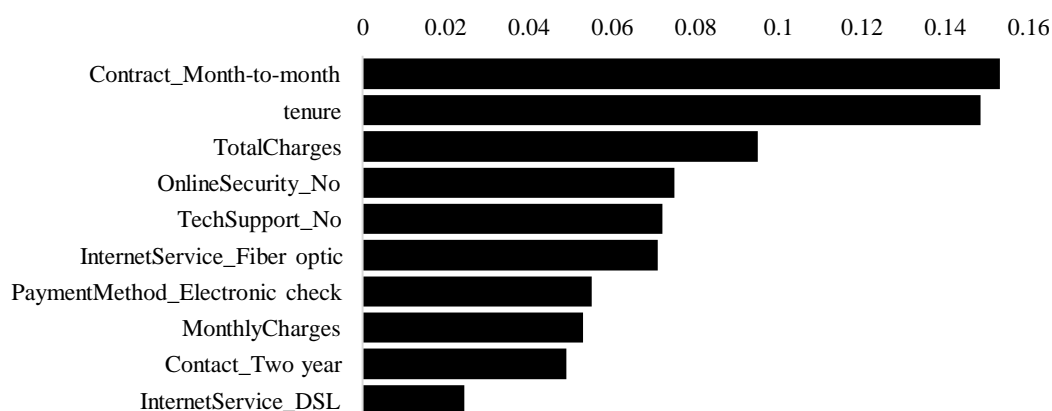
Sources: developed by the authors.

Let us single out certain observations:

- Subscription for a 2-year contract decreases the likelihood of outflow. Tenure and a 2-year contract have the most negative relationship with outflow, as predicted by modelling with logistic regression.
- Subscription for the service of DSL Internet also lessens churn.
- General costs, internet services, monthly contracts, and the age of users can lead to an increase in churn rate. This is quite curious because even though fibre optic services have higher speeds than others, subscribers are more likely to stop using the services because of this.



Implementing the Random Forest model helps achieve accuracy at 80.9%, which is quite similar to logistic regression. Based on the algorithm of Random Forest, the tenure, monthly contract, and general cost are the most important forecast characteristics for churn prediction and modelling. The results of this algorithm are similar to the results of logistic regression.



**Figure 8. The importance of factors which influenced customer churn, determined by the Random Forest model**

Sources: developed by the authors.

Implementation of the SVM algorithm helps to increase accuracy to 82%. Analysis of the confusion matrix (Table 2) shows that false negative rates are lower than in the logistic regression model (8,5% vs 10,3%).

**Table 2. Confusion matrix from SVM algorithm**

True label	Predicted label	
	953	89
	164	201

Sources: developed by the authors.

At the final stage, we construct the XGBoost model based on boosting algorithms, which helps to increase accuracy to 83%. After using various Data Science models, the best prediction model for this dataset was XG Boost, with the highest accuracy. However, the algorithm of support vector machines gives the lowest level of false-negative results. As a result, the combination of XG Boost and SVM is recommended for implementation in a telecommunications company to predict the churn rate for the customer base.

**Conclusions.** The active implementation of big data has become a requirement in the development process and business decisions in the 21st century, thanks to the ever-increasing opportunity to produce and collect it at an extraordinary speed. Thus, the categories «data analytics», «big data», and «Data Science» have become frequently used in the last 5-10 decades, which caused their active application in numerous areas of economics and social life.

Reducing the outflow of customers is one of the most important tasks of the marketing activity of a telecommunications enterprise operating in a highly competitive market. To guarantee the effective development of business, there is a need to use Data Science and machine learning technologies.

It was determined that methods of BDA began to show more and more applications in practice for business, in particular in the telecommunications industry, in order to evaluate the level of clients' satisfaction with company services, determine factors that cause dissatisfaction, and foresee which clients are at a greater risk of abandoning and changing services supplier.

Data Science and machine learning allow management and the marketing department to establish marketing processes dedicated to retaining customers who tend to switch suppliers. The study proves the necessity of the implementation of Data Science models and machine learning for the telecommunications market because they demonstrate high-quality forecasting of key indicators and make it possible to improve business results.

At the modelling stage, the main models (Random Forest, XGBoost, logistic regression, and support vector machines) were implemented using the Python programming language. The assessment of the accuracy of the results reached more than 80%, which points out the possibility and feasibility of models' applications to predict those who may abandon the company's services and minimise the outflow of customers based on this.

The built models are effective because they have high accuracy and a sufficiently low level of false positive/negative results (within 10%). A high-quality classification of client loyalty allows one to adequately determine the rate of client loyalty and the intention to leave the provider with a general accuracy of more than 80%. As the results of the study are planned to be implemented to prevent customer outflow when planning marketing activities, it is appropriate to prioritise the models considering both accuracy and easiness of interpretation.

The key factors affecting the churn of clients from the company's client base are the type of contract (monthly or long-term), the duration of using the services, and the costs of the telecommunication company's services. Managing these factors and finding additional motives for managing customer behaviour is the basis for ways to improve the company's services and increase the business's operational results. As the proposed models built in the study showed high accuracy in the classification of customer loyalty, i.e. the client's desire to stay with the enterprise or switch to another telecommunication provider, there is an opportunity to implement proactive management promptly to prevent customer churn (since acquiring a new client is from 5 to 10 times more expensive than keeping a current one client), retaining clients satisfaction and loyalty, and reducing expenses.

The results of the research can be useful for optimising the marketing activity of managing the outflow of consumers of enterprises in the market of telecommunications by making effective decisions based on data analysis and modelling and as well as for improving the mathematical methodology of forecasting the outflow of consumers. Therefore, the most significant theoretical and practical inferences of the study are to develop an efficient forecasting system for businesses to monitor churn risks and to enrich the scientific resources on data analytics, machine learning methods, and methodology of Data Science to identify vital factors that determine the propensity of customers to churn. Additionally, regular control is relevant for the timely identification of new potential factors that will lead to customer churn. Furthermore, it is recommended to carry out research periodically and consider new additional variables which could influence the churn rate.

The main limitations of this study include only four machine learning models and sample imbalance, which can be surmounted by implementing additional balancing procedures and applying particularly price-sensitive classification methods. Models may become outdated due to fundamental changes in market conditions or tariff plans, a decline in the service and quality of communication, a substantial change in positions on the market, or large changes in the economic or political situation in the country. To prevent such a situation, enhancing the models monthly and reacting to possible threats in time is relevant.

**Author Contributions:** conceptualisation, T. Z and Y. F; methodology, T. Z and Y. F; software, Y. F; validation, T. Z and Y. F; formal analysis, V. S and Y. F; investigation, T. Z and V. S; resources, T. Z and Y. F; data curation, V. S and Y. F; writing-original draft preparation, T. Z and Y. F; writing-review and editing, T. Z and Y. F; visualisation, Y. F; supervision, T. Z; project administration, T. Z.

## References

- Arya, K., Kumar, T., & Jain, M. K. (2016). Big data analytics of global E-commerce organisations: A study survey and analysis. *Int. J. Sci. Eng. Res.*, 7(12), 82-84. [\[Google Scholar\]](#)
- Chornous, G., & Farenjuk, Y. (2021). Marketing mix modelling for pharmaceutical companies on the basis of data science technologies. *Access to science, business, innovation in digital economy*, 274-289. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Dai, X. (2017). Identifying dissatisfied 4G customers from network indicators: a comparison between complaint and survey data. In *Big Data Applications in the Telecommunications Industry* (pp. 41-53). IGI Global. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Fang, X. (2021, August). Research on digital marketing strategy of telecommunication service based on computer complex network model. In *Journal of Physics: Conference Series* (Vol. 1992, No. 4, p. 042002). IOP Publishing. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Farenjuk, Y. (2022). Marketing strategy optimisation in fmcg market. *Bulletin of Taras Shevchenko National University of Kyiv. Economics*, 2(219), 49-57. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Fedirko, O., Zatonatska, T., Dlihopolskyi, O., & Londar, S. (2019). The impact of e-commerce on the sustainable development: case of Ukraine, Poland, and Austria, *IOP Conference Series Earth and Environmental Science*, 915(1), 012023. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Goworek, K. (2021). The big impact of Big Data on the telecom industry. [\[Google Scholar\]](#) [\[CrossRef\]](#)

- Hsu, K. K. (2019). Discussion on the live broadcast of social media and e-commerce. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Hurtado, P. A., Dorneles, C., & Frazzon, E. (2019). Big Data Application for E-commerce's Logistics: A research assessment and conceptual model. *IFAC-papers online*, 52(13), 838-843. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Maroufkhani, P., Wagner, R., Wan Ismail, W. K., Baroto, M. B., & Nourani, M. (2019). Big data analytics and firm performance: A systematic review. *Information*, 10(7), 226. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Moorthi, K., Dhiman, G., Arulprakash, P., Suresh, C., & Srihari, K. (2021). Withdrawn: A survey on the impact of data analytics techniques in E-commerce. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Parveen, N. (2021). Sentiment Analysis on Big Data Using Machine Learning Algorithms. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(9), 47-52. [\[Google Scholar\]](#)
- Radukic, S., Mastilo, Z., & Kostic, Z. (2019). Effects of digital transformation and network externalities in the telecommunication markets. *Economics*, 7(2), 31-42. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Riddle, J. (2020). How Will Big Data Transform E-Commerce Marketplaces? [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Rosario, A., & Raimundo, R. (2021). Consumer marketing strategy and E-commerce in the last decade: a literature review. *Journal of Theoretical and applied electronic commerce research*, 16(7), 3003-3024. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Ryfiak, S. (2020). Big Data is taking eCommerce by storm. Here's why you can't wait it out. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Sekli, G. F., & Vega, I. (2021). Adoption of Big Data analytics and its impact on organizational performance in higher education mediated by knowledge management. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(4), 221. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Simaković, M. N., Cica, Z. G., & Masnikosa, I. B. (2021). *Big Data architecture for mobile network operators*. 15th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), 283-286. [\[CrossRef\]](#)
- Singh, R. K., & Verma, H. K. (2020). Influence of social media analytics on online food delivery systems. *International Journal of Information System Modeling and Design (IJISMD)*, 11(3), 1-21. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Thakkar, H. K., Desai, A., Ghosh, S., Singh, P., & Sharma, G. (2022). Clairvoyant: AdaBoost with cost-enabled cost-sensitive classifier for customer churn prediction. *Computational Intelligence and Neuroscience*, 2022. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Gabor, M. R., & Varga, I. E. (2021) the influence of social networks in travel decisions. *Economics-časopis za inovacijska i ekonomska istraživanja*, 9(2). [\[Google Scholar\]](#)
- Yang, M., Ren, Y., & Adomavicius, G. (2019). Understanding user-generated content and customer engagement on Facebook business pages. *Information Systems Research*, 30(3), 839-855. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Yen, Y. S. (2014). The interaction effect on customer purchase intention in e-commerce: A comparison between substitute and complement. *Asia Pacific Journal of Marketing and Logistics*, 26(3), 472-493. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Yusuf-Asaju, A. W., Dahalin, Z. B., & Ta'a, A. (2017, May). Mobile network quality of experience using big data analytics approach. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 658-664). IEEE. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Zhang, D., Pee, L. G., & Cui, L. (2021). Artificial intelligence in E-commerce fulfilment: A case study of resource orchestration at Alibaba's Smart Warehouse. *International Journal of Information Management*, 57, 102304. [\[Google Scholar\]](#) [\[CrossRef\]](#)

**Тетяна Затонацька**, д.е.н., професор, Київський національний університет імені Тараса Шевченка Україна  
**Яна Фаренюк**, Київський національний університет імені Тараса Шевченка, Україна  
**Віктор Шпирко**, к.е.н., Київський національний університет імені Тараса Шевченка, Україна  
**Модельовання рівня відтоку для телекомунікаційних компаній з використанням методів Data Science**

Телекомунікаційні компанії функціонують на ринку з надзвичайно високою конкуренцією. Залучення нового клієнта в 5-10 разів дорожче, ніж утримання існуючого. Таким чином, ефективне управління відтоком абонентів і розуміння його причин стали досить актуальними завданнями для операторів мобільного зв'язку. У зв'язку з цим прогнозування відтоку абонентів через активність конкурентів стає дуже важливим. Data Science і машинне навчання створюють широкі можливості для вирішення цього завдання, щоб оцінити рівень задоволеності клієнтів послугами компанії, виявити фактори, що викликають незадоволення, і передбачити, які клієнти мають більший ризик відмовитися від послуг і змінити постачальника. Компанія, яка впроваджує аналіз даних і

моделювання для розробки моделей прогнозування відтоку клієнтів, має можливість покращити управління відтоком клієнтів і підвищити результати бізнесу. Метою дослідження є застосування моделей машинного навчання для телекомунікаційної компанії, зокрема побудова моделей для прогнозування відтоку користувачів і доведення того, що моделі Data Science і машинне навчання є якісними та ефективними інструментами для вирішення завдань прогнозування ключових маркетингових показників телекомунікаційної компанії. На прикладі Telco у статті представлено результати різних моделей класифікації, таких як логістична регресія, Random Forest, SVM та XGBoost з використанням мови програмування Python. Всі моделі мають високий рівень якості (загальна точність більше 80%). Таким чином, дослідження доводить можливість та доцільність використання моделей у подальшій класифікації клієнтів для прогнозування відтоку абонентів (тих клієнтів, які можуть відмовитися від послуг компанії) та мінімізації відтоку споживачів на основі цього. Виявлено ключові фактори, що впливають на відтік клієнтів, і вони створюють основу для майбутнього прогнозу відтоку. Впровадження моделей прогнозування відтоку абонентів допоможе зменшити його і зберегти їх лояльність. Результати дослідження можуть бути використані для оптимізації маркетингової діяльності з управління відтоком споживачів компаній ринку телекомунікацій шляхом прийняття ефективних рішень на основі даних та для вдосконалення математичної методології прогнозування відтоку клієнтів. Таким чином, основні теоретичні та практичні наслідки дослідження полягають у розробці ефективного інструменту прогнозування для менеджерів для контролю ризиків відтоку та збагаченні літератури з аналізу даних і моделей Data Science для визначення критичних факторів, які визначають схильність клієнтів до відтоку.

**Ключові слова:** відтік, споживач, Data Science, прогнозування, клієнт, машинне навчання, маркетинг, моделювання.