A Review on Machine Learning-Based Customer Churn Prediction in the Telecom Industry

Sawsan Barham, Nowfal Aweisi, and Ala' Khalifeh, *Member, IEEE*

2023 9th International Conference on Control, Decision and Information Technologies (CoDIT) | 979-8-3503-1140-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/CODIT58514.2023.10284430

***Abstract*— In today's business landscape, companies are facing significant difficulties in achieving positive client interactions and maintaining customers' satisfaction. As a result, businesses are striving to focus on every aspect related to customers and their behaviors to compete in the industry. This has become increasingly important for building customers' loyalties, given the many opportunities available to customize services or products for each customer. To prevent customers from becoming dissatisfied and leaving, businesses are deploying various techniques to efficiently predict customers' behaviors and identify those who may churn or stop using the company's services or products. The rise of machine learning applications has significantly contributed to addressing the challenge of predicting customers' churn rates. Researchers worldwide are now moving towards applying machine learning techniques in this area. This paper aims to present a review of various studies conducted from 2019 to 2022 that utilized machine-learning techniques to predict customers' churn in the telecom industry. The paper summarizes the different machine learning algorithms that are used for customers' churn prediction, with a particular focus on the telecom industry, as well as their accuracy, to provide insights into the effectiveness of these techniques in addressing customers' churn in the telecom industry.**

I. Introduction

The success of any business relies on the satisfaction of its customers, which is generally measured by how happy they are with the provided products or services. Therefore, organizations aim to achieve high levels of customers' satisfaction by offering top-quality services and staying attuned to the ever-changing needs of a dynamic market. This approach has led to the emergence of the concept of customers' retention, which is the ability of a business to maintain a loyal customer base over time by keeping customers satisfied. Customers' retention is vital for businesses in several ways. Firstly, it can be more cost-effective than acquiring new customers, as it can be more expensive to attract new customers than to retain existing ones. Secondly, loyal customers are more likely to recommend a business to others, which can lead to increased sales and revenue. Finally, retaining customers over the long-term can help a business build a strong brand reputation and establish itself as a trusted and reliable provider of products or services. Therefore, achieving high levels of customers' satisfaction and implementing effective customer retention strategies are critical for the success and growth of any business. "A customer retention rate is an important key performance indicator (KPI) for companies, especially from the customer relationship management (CRM) point of view. It expresses how much the company succeeds in maintaining a long-term relationship with its customers." [1]. another definition: "Customer retention is a statistic that gauges customer loyalty and an organization's ability to maintain customers over time. Consumer retention can indicate or forecast customer happiness, repurchase behavior, customer involvement, and emotional links to a brand, in addition to recognizing the number of loyal customers." [2]. In today's competitive business landscape, companies must be proactive in customer retention to avoid costly losses and maintain a unique market position. Churn analysis is a tool that helps predict customers' churn and implement proactive measures. The telecom industry presents unique challenges in analyzing diverse customers' data, making it difficult to identify patterns and trends. Despite these challenges, many companies invest in churn analysis tools to improve customers' retention rates and maintain their market position. Interpretable data and an accurate analysis of customers' behavior are crucial for a successful churn analysis.

A. Telecoms Sector, Churn Analysis is Critical.

Communication demands are now frequently regarded as a need rather than an added cost. With each passing day, the market's growth pace slows down. Gaining new subscribers is getting increasingly tricky, considering these reviews. As a result, businesses conduct actions such as churn analysis and activities to maintain customers in the system. It is clear why churn analysis is crucial in the telecommunications sector when considering the following factors:

- For telecoms firm with churning subscribers each year, quarterly turnover is a hidden cost.
- Continuity costs are five times lower than new client acquisition costs.

Companies aim to maximize profits by retaining customers for as long as possible. Long-term contracts and tariffs are preferred over short-term relationships, as they provide a more stable revenue stream. To achieve this goal, companies use churn analysis to identify customers who are at risk of leaving and take steps to retain them. By cultivating customers' loyalty, companies can improve customers' retention rates and boost their profitability over time. Therefore, to assist the industry in developing such tools and strategies, this paper provides a survey of the recent work that utilizes machine learning algorithms to predict churning customers, with a focus on the telecom sector being one of the most crucial industries, thus giving companies a heads-up toward improving their services and giving these customers extra care. The rest of the paper is organized as follows.

Sawsan Barham, Princess Sumaya University for Technology, Amman, Jordan (e-mail: saw20208124@std.psut.edu.jo).
Nowfal Aweisi, Princess Sumaya University for Technology, Amman, Jordan (e-mail: now20208037@std.psut.edu.jo).
Ala' Khalifeh is with German Jordanian University, Amman, Jordan (e-mail: ala.khalifeh@gju.edu.jo).

Section II provides a literature review on similar work papers that address customers' churn prediction. The methodology that is adopted in this paper is described in section III. Then, section IV presents the paper's results and discussion. Finally, section V concludes the paper and proposes future work.

II. Literature review

Different papers discussed customers' churn prediction from various aspects [1]. However, few papers provided a review of customers' churn prediction utilizing machine learning (ML) algorithms. Furthermore, although there has been significant research work on predicting telecom, customers' churn rate, however, there is a lack of surveying these papers and providing a summary of the used machine learning models, the utilized datasets, and the customers' churn prediction accuracy. In the literature, De et al. studied various models utilized in existing papers and provided a summary of the ML techniques used for customers' churn prediction. Fujo et al. [3] implemented a model that predicts telecom customers' churn using different ML techniques. Results showed that they were effective and accurate. Different papers discussed customers' churn prediction from various aspects, such as the work presented in [4], which presented a systematic review in which the authors estimated the evolution of actual studies about the prediction of customers' dropout using ML. However, the conducted literature review shows that there is a need for a survey paper that summarizes the work related to customers' churn prediction with an emphasis on the telecom industry, given the fact that there are several datasets available in this domain and several papers that are focused on customers' churn prediction using ML.

III. Research Methodology

To identify the best related papers to this work, a thorough search was conducted on Google Scholar and in the major scientific databases. The search targeted papers related to customers' churn prediction that were published between 2019- 2022. Several keywords related to this review were used. A combination of phrase search and keywords search was used such as: customer retention churn, customer retention, churn prediction, telecommunication, telecommunication industries, machine learning, machine learning techniques, customer satisfaction, gaining customer loyalty, etc. The initial search results were about 1370 papers, which were decreased to 512 through narrowing the search criteria using the search keywords, to reach a specific set of studies that best match the search requirement and can answer the review questions. Furthermore, the search was narrowed further until 120, and the results were screened by dividing the studies into two groups and reviewed through title and abstract screening based on the exclusion and inclusion criteria, then the results were screened again through a full-text screening to include studies that comply better with the inclusion criteria. The final screening results identified 33 relevant studies to be included in this review that contribute to the research objectives.

IV. Results and Discussion

To have a comprehensive overview of customers' churn prediction, a summary of the most related works in the telecom industry and in other fields as well is provided. Therefore, Table 1 focused on the telecom industry, while Table II summarizes other papers related to other industries such as e-Commerce, finance, health, etc. Both tables provided a summary of the most crucial information needed to assess the paper's relevance, the utilized dataset, and the ML models used with their reported prediction accuracy.

The reviewed papers showed that many articles used telecom datasets for research, with nine using IBM Inc. telco data and three using University of California- Irvine (UCI) datasets. The accuracy results varied between studies, highlighting the need for careful consideration of factors affecting churn prediction models. The study used various datasets, including bank, retail, hospitality, and marketing agency datasets. The primary focus was on 2021 and 2022 surveys. A high number of articles used Decision Tree, Random Forest, Logistic Regression, Naïve Bayes, Neural Network, and XGBoost classifiers. Some researchers used multiple classifiers simultaneously, thus enhancing the accuracy. Furthermore, the third observation is that some researchers have used a combination of two or more classifiers simultaneously, which resulted in an enhancement in the accuracy result [5].

From the surveyed papers as mentioned in Table I, it has been found that the Random Forest algorithm has achieved the best performance in the telecom customers' churn prediction in most of the surveyed studies, where the following papers [6, 8, 9, 16] reported relatively high prediction rates of 91.66%, 97.4%, 87.7%, and 98%, respectively, while the following papers [11, 12, 13, 20] have reported medium accuracy of 79%, 78%, 81%, and 79%, respectively. The Decision Tree algorithms have achieved the second highest customers' churn prediction accuracy in the surveyed papers, where the following papers [6, 7, 10, 16] reported high accuracy results of 90.9%, 86.1%, 94.98% and 90%, respectively, while the following papers [9, 11, 12] have reported medium accuracy of 83%, 78%, and 80.14%, respectively. On the other hand, the Logistic Regression achieved the third highest customers' churn prediction accuracy of 93.94%, and 87% in the papers [25, 7], respectively, While the following papers [11, 12, 13, 20, 24] have reported medium accuracy of 81%, 80%, 80.29%, 79.5%, and 80.48%, respectively. Furthermore, the following algorithms have been used less frequently in customers' churn prediction compared with the mentioned algorithms, where the Native Bayes algorithm achieved relatively high-performance prediction in paper [6], where 86.53% prediction accuracy has been reported, while the algorithm had medium prediction accuracy in [19, 12], with 82% and 77.07%, respectively. Finally, the XGBoost algorithm achieved a high prediction performance of 93.3% in this paper [9], while it achieved a medium prediction performance of 80.81%, 82.2%, 80.7%, 77%, and 79.35% in the following papers [12, 13, 20, 22, 24], respectively.

Furthermore, as mentioned in Table II, it has been found that in other industries, the Random Forests and the Logistic Regression algorithms have been mostly used in the surveyed papers, followed by the Decision Tree and the Naive Bayes algorithms. Thus, recommending them for customers' churn prediction in the telecom and other industries.

V. CONCLUSION AND FUTURE WORK

To conclude, customers' churn rates are considered an indicator of customers' satisfaction. A low dropout rate suggests that customers are happy, while high rates indicate that customers are leaving due to dissatisfaction. The purpose of this review paper is to provide an update on the current state of research that utilizes machine learning to predict churning customers, specifically between the years 2019 and 2022. The accuracy of various machine learning techniques used for predicting customers' churn rates was summarized, and information from other studies was extracted regarding the methods most commonly used in predicting customers' churn rates. In a future work, more analysis should be conducted to identify other factors that affect churn prediction accuracy.

REFERENCES

[1] T. W. Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, D. Ratj, and B. Pardamean, "Deep learning as a vector embedding model for customer churn," Procedia Computer Science, vol. 179, pp. 624–631, 2021, 5th International Conference on Computer Science and Computational Intelligence 2020.

[2] S. De, P. P, and J. Paulose, "Effective ml techniques to predict customer churn," in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 895–902.

[3] S. Wael Fujo, S. Subramanian, and M. Ahmad Khder, "Customer churn prediction in telecommunication industry using deep learning," Information Sciences Letters, vol. 11, pp. 185–198, 2022.

[4] P. Sobreiro, D. D. S. Martinho, J. G. Alonso, and J. Berrocal, "A slr on customer dropout prediction," IEEE Access, vol. 10, pp. 14 529–14 547, 2022.

[5] M. J. Shabankareh, M. A. Shabankareh, A. Nazarian, A. Ranjbaran, and N. Seyyedamiri, "A stacking-based data mining solution to customer churn prediction," Journal of Relationship Marketing, vol. 21, no. 2, pp. 124–147, 2022.

[6] A. Mishra and U. S. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble-based classifiers," in 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, pp. 721–725.

[7] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," European Journal of Operational Research, vol. 269, no. 2, pp. 760–772, 2018.

[8] H. Faris, "A hybrid swarm intelligent neural network model for customer churn prediction and identifying the influencing factors," Information, vol. 9, no. 11, 2018.

[9] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 28, 2019.

[10] A. S´niegula, A. Poniszewska-Maran´da, and M. Popovic´, "Study of machine learning methods for customer churn prediction in telecommunication company," in Proceedings of the 21st International Conference on Information Integration and Web Based Applications and amp Services, ser. iiWAS2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 640–644.

[11] M. Pelka and A. Rybicka, "Hybrid conjoint analysis – symbolic decision tree model for customer churn prediction model," in 34th IBIMA Conference: 13-14 November 2019, Madrid, Spain. Madrid, Spain: IBIMA, 2019, pp. 12 435–12 441.

[12] A. C¸ Alıs¸, and J. Kozłowska, "Customer churn prediction with popular machine learning algorithms," Akademia Zarzadzania, 01, 2021.

[13] P. Lalwani, M. Mishra, J. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," Computing, vol. 104, pp. 271–294, 02 2022.

[14] L. Hota and P. K. Dash, "Prediction of customer churn in telecom industry: A machine learning perspective," Computational Intelligence and Machine Learning, vol. 2, 10 2021.

[15] N. Mustafa, L. Sook Ling, and S. Abdul Razak, "Customer churn prediction for telecommunication industry: A Malaysian case study]," F1000Research, vol. 10, no. 1274, 2021.

[16] B. Saputro, S. Ma'mun, I. Budi, A. B. Santoso, and P. K. Putra, "Customer churn factors detection in Indonesian postpaid telecommunication services as a supporting medium for preventing waste of IT components," IOP Conference Series: Earth and Environmental Science, vol. 700, no. 1, p. 012015, 03 2021.

[17] B. Sarkar, R. Islam, and M. S. Hossain, "Customer churn prediction for telecommunication operator," Daffodil International University, Dhaka, Bangladesh, Tech. Rep., 2021.

[18] A. Dalli, "Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector," Mathematical Problems in Engineering, vol. 2022, no. 4720539, 02 2022.

[19] M. Fraihat, S. Fraihat, M. Awad, and M. AlKasassbeh, "An efficient enhanced k-means clustering algorithm for best offer prediction in telecom," International Journal of Electrical and Computer Engineering (IJECE), vol. 12, no. 3, pp. 2931–2943, 2022.

[20] D. Geetha, P. Tomar, and A. Jain, "Customer churn prediction," Turkish Journal of Computer and Mathematics Education, vol. 13, no. 3, pp. 944–948, 2022.

[21] T. Kimura, "Customer churn prediction with hybrid resampling and ensemble learning," Journal of Management Information and Decision Sciences, vol. 25, pp. 1–23, 01 2022.

[22] R. H. Mohammed Hend, Osama Emam, "The role of decision support system in enhancing customer relation management in the Egyptian telecommunication sector," IEEE-SEM, vol. 10, 2022.

[23] J. Rabbah, M. Ridouani, and L. Hassouni, "A new churn prediction model based on deep insight features transformation for convolution neural network architecture and stacknet," International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), vol. 17, no. 1, pp. 1–18, 2022.

[24] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "Data transformation based optimized customer churn prediction model for the telecommunication industry," ArXiv, vol. abs/2201.04088, 2022.

[25] Y. Singh, Y. Pandit, and N. Joshi, "Prediction of customer churn using machine learning," International Research Journal of Modernization in Engineering Technology and Science, vol. 04, 2022.

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

[27] I. Ullah, H. Hussain, S. Rahman, A. Rahman, M. Shabir, N. Ullah, and K. Ullah, "Using k-means, lof, and cblof as prediction tools," International Journal of Computational Intelligence in Control, vol. 13, 06 2021.

[28] P. Routh, A. Roy, and J. Meyer, "Estimating customer churn under competing risks," Journal of the Operational Research Society, vol. 72, no. 5, pp. 1138–1155, 2021.

[29] N. T. M. Sagala and S. D. Permai, "Enhanced churn prediction model with boosted trees algorithms in the banking sector," in 2021 International Conference on Data Science and Its Applications (ICoDSA), 2021, pp. 240–245.

[30] O. F. Seymen, O. Dogan, and A. Hiziroglu, "Customer churn prediction using deep learning," in Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). Cham: Springer International Publishing, 2021, pp. 520–529.

[31] B. N. Bristy, "Customer churn analysis and prediction," master's thesis, United International University, Dhaka, Bangladesh, 2022.

[32] S. Kim and H. Lee, "Customer churn prediction in influencer commerce: An application of decision trees," Procedia Computer Science, vol. 199, pp. 1332–1339, 2022.

[33] K. Matuszelan´ski and K. Kopczewska, "Customer churn in retail e-commerce business: Spatial and machine learning approach," Journal of Theoretical and Applied Electronic Commerce Research, vol. 17, no. 1, pp. 165–198, 2022.

Table 1: A summary table of the surveyed papers in customers' churn prediction in the telecom industry

Authors	Title	Year	Techniques	Accuracy	Dataset	Country	Ref
A. Mishra & U. S. Reddy	A Comparative Study of Customer Churn Prediction In Telecom Industry Using Ensemble-based Classifiers	2017	Bagging Boosting Random Forest Decision Tree Support Vector Machines Naive Bayes	90.83% 90.32% 91.66% 90.9% 90.12% 86.53%	The University of California, Irvine (3333 customers)	India	[6]
A. De Caigny & K. Coussement & K. W. De Bock	A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees	2018	Logit Leaf Model Decision Tree Logistic Model Tree Logistic Regression Random Forest	62-88.7% 60.1-86.1% 61.8-88.5% 59.5-87.9% 61.4-87.7%	7 datasets financial services (2742849 entries) 3 datasets telecom (168808 entries) 1 datasets retail (32371 entries) 1 dataset DIY (3827 entries) 1 dataset newspaper (427833 entries) 1 dataset energy (20000 entries)	France	[7]
H. Faris	A Hybrid Swarm Intelligent Neural Network Model for Customer Churn Prediction and Identifying The Influencing Factors	2018	ADASYN-PSO-wNN PSO-wNN Bagging AdaBoost Random Forest Grid-Support Vector Machines MLP (17-10-2) MLP(17-10-10-2)	Jordan-USA 96.3 - 92% 92.2 - 92.3% 97.6 - 95.3% 97 - 95.2% 97.4 - 95.9% 92.4 - 93.1% 92.16 - 94.6% 92.27 - 94.9%	DKD Dataset US mobile operator (3333 customers) Major cellular telecommunication company in Jordan (5000 customers)	Jordan	[8]
A. K. Ahmad & A. Jafar & K. Aljoumaa	Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform	2019	XGBoost Gradient Boosted Machine Random Forest Decision Tree	93.3% 90.89% 87.76% 83%	70 Terabyte SyriaTel	Syria	[9]
A. S´niegula & A. Poniszewska-Maran´da & M. Popovic	Study of Machine Learning Methods for Customer Churn Prediction In Telecommunication Company	2019	K-Means Decision tree Artificial Neural Networks	62.65% 94.98% 87.11%	Bigml database (3333 customers)	Poland	[10]
M. Pelka & A. Rybicka	Hybrid Conjoint Analysis – Symbolic Decision Tree Model for Customer Churn Prediction Model	2019	Symbolic Decision Tree	--------	Poland (109 customers)	Poland	[11]
A. C¸ Alıs & J. Kozłowska	Customer Churn Prediction with Popular Machine Learning Algorithms	2021	Logistic Regression K-Nearest Neighbor Decision Tree Random Forest Support Vector Classifier	81% 76% 78% 79% 78%	California (7043 customers)	Turkey	[12]
P. Lalwani & M. Mishra & J. Chadha & P. Sethi	Customer Churn Prediction System: A Machine Learning Approach	2021	Logistic Regression LR (Adaboost) Decision Tree Adaboost Adaboost (Extra Tree) K-Nearest Neighbor Random Forest RF (Adaboost) Naive Bayes XGBoost CatBoost Support Vector Machines SVM Linear SVM Poly SVM (Adaboost)	80.45% 76.57% 80.14% 81.71% 81.14% 79.64% 78.04% 81.21% 77.07% 80.8% 81.8% 79.14% 80.21% 74.07%	Brazil (around 7000 customers)	India	[13]
L. Hota & P. K. Dash	Prediction of Customer Churn in Telecom Industry: A Machine Learning Perspective	2021	GradientBoost AdaBoost XGBoost Artificial Neural Networks Logistic Regression Random Forest	80.41% 80.59% 82.2% 79.98% 80.29% 81.10%	IBM Telecom's Kaggle Dataset (7043 customers)	India	[14]
M. J. Shabankareh, & M. A. Shabankareh & A. Nazarian & A. Ranjbaran & N. Seyyedamiri	A Stacking-based Data Mining Solution to Customer Churn Prediction	2021	Support Vector Machines SVM + CHAID SVM + C5 SVM + C&R Tree SVM + Neural Net SVM + K-Nearest Neighbor	 85.8% 85.4% 84.9% 82.8% 81.2%	7043 customers	Iran	[5]

N. Mustafa, & L. Sook Ling & S. Abdul Razak	Customer Churn Prediction for Telecommunication Industry: A Malaysian Case Study	2021	Logistic Regression Linear Discriminant Analysis K-Nearest Neighbor Classification and Regression Trees Naive Bayes Support Vector Machines	41% 42% 98% 98% 41% 98%	Malaysian telecommunication company (7776 records)	Malaysia	[15]
B. Saputro & S. Ma'mun & I. Budi & A. B. Santoso & P. K. Putra	Customer Churn Factors Detection in Indonesian Postpaid Telecommunication Services As A Supporting Medium for Preventing Waste of IT Components	2021	Fast Large Margin Support Vector Machines Naive Bayes	71.4% 65% 60.8%	16.626 customers	Indonesia	[16]
B. Sarkar & R. Islam & M. S. Hossain	Customer Churn Prediction for Telecommunication Operator	2021	Adaboost Decision Tree Bagging Random Forest	97% 90% 98% 98%	The Iranian telecommunication company (3150 records)	Bangladesh	[17]
T. W. Cenggoro & R. A. Wirastari & E. Rudianto & M. I. Mohadi & D. Ratj & B. Pardamean	Deep Learning as A Vector Embedding Model for Customer Churn	2021	Vector embedding in Deep Learning	89.82%	Kaggle" India" publicly available dataset for telecommunication customer churn (3333 customers)	Indonesia	[1]
A. Dalli	Impact of Hyperparameters on Deep Learning Model for Customer Churn Prediction in Telecommunication Sector	2022	Neural Network Model	86.9%	Turkish telecommunication Company Dataset (3333 customers)	Morocco	[18]
M. Fraihat & S. Fraihat & M. Awad & M. AlKasassbeh	An Efficient Enhanced K-means Clustering Algorithm for Best Offer Prediction in Telecom	2022	Classical K-Means Enhanced K-Means	<70% >90%	CDRs of a telecom company in Jordan (100000 customers)	UAE	[19]
D. Geetha & P. Tomar & A. Jain	Customer Churn Prediction	2022	Adaboost Naive Bayes	84% 82%	Kaggle dataset called Telecom Customer Churn 7043 customers	India	[20]
T. Kimura	Customer Churn Prediction With Hybrid Resampling and Ensemble Learning	2022	Logistic Regression Support Vector Machines Random Forest XGBoost LightGBM CatBoost	79.5% 78.4% 79.1% 80.7% 79.5% 79.8%	IBM Telco 7,043 Customers	Japan	[21]
M. Hend & O. Emam & R. Haggag	The Role of Decision Support System in Enhancing Customer Relation Management in The Egyptian Telecommunication Sector	2022	Neural Network Multilayer Perceptron Neural Network Voted Perceptron Decision Tree J48 Logistic Model Tree Decision Rule PART Decision Rule JRip Naive Bayes Naive Bayes updateable	94.65% 84.78% 95.44% 96.47% 95.19% 96.22% 90.35% 90.35%	cell2cell (3333 customers)	Egypt	[22]
J. Rabbah & M. Ridouani & L. Hassouni	A New Churn Prediction Model Based on Deep Insight Features Transformation for Convolution Neural Network Architecture and Stacknet	2022	StackNet Model Decision Tree Logistic Regression Random Forest XGBoost Naive Bayes	83% 80% 73% 75% 77% 84%	California (7043 customers)	Morocco	[23]
J. K. Sana & M. Z. Abedin & M. Sohel Rahman & M. Saifur Rahman	Data Transformation Based Optimized Customer Churn Prediction Model For The Telecommunication Industry	2022	K-Nearest Neighbor Naive Bayes Random Forest Logistic Regression Feed-Forward Neural Networks (FNN) Recurrent Neural Networks Decision Tree (DT) Gradient Boosting (GB)	Dataset 1 FNN 80.2% Dataset 2 DT & GB >82% Dataset 3 DT 84% GB 86.4%	Dataset 1: Kaggle (100000 entries) Dataset 2: The University of California, Irvine (5000 entries) Dataset 3:Kaggle(3333 customers)	Bangladesh	[24]
S. Wael Fujo & S. Subramanian & M. Ahmad Khder	Customer Churn Prediction in Telecommunication Industry Using Deep Learning	2022	Deep-BP-ANN	IBM 88.12% Cell2Cell 79.38%	IBM Telco. (7044 customers) Cell2Cell (71047 customers)	Bahrain	[3]
Y. Singh & Y. Pandit & N. Joshi	Prediction of Customer Churn Using Machine Learning	2022	Random Forest Gaussian Naive Bayes K-Nearest Neighbor Neural Network Extra Tree Logistic Regression XGBoost LightGBM	77.41% 73.89% 76.39% 78.15% 76.39% 80.48% 79.35% 79%	Telecom Customer Churn dataset, which is available on Kaggle (7043 customers)	India	[25]

T. Zhang & S. Moro & R. F. Ramos	A Data-driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation	2022	Fisher Linear Discriminant Equation Logistic Regression	78% 93.94%	Three major Chinese telecom companies China Mobile, China Unicom, and China Telecom. 4126	Portugal	[26]

Table II: A summary table of the surveyed papers in customers' churn prediction in other industries

Authors	Title	Year	Techniques	Accuracy	Dataset/Industry	Country	Ref
I. Ullah & H. Hussain & S. Rahman & A. Rahman & M. Shabir & N. Ullah & K. Ullah	Using K-Means, LOF, and CBLOF As Prediction Tools	2021	K-Means (Breast Cancer) K-Means (Banking) Local Outlier Factor (Breast Cancer) Local Outlier Factor (Banking) Clusters-Based-Local-Outliers Factors (Breast Cancer) Clusters-Based-Local-Outliers Factors (Banking)	76.95% 74.36% 78.36% 68% 84% 80.38%	Pakistan Bank (4521 customers)	Pakistan	[27]
P. Routh & A. Roy & J. Meyer	Estimating Customer Churn Under Competing Risks	2021	Competing Risk Random Survival Forest	Improves accuracy up to 20%	Hospitality industry. (1368 customers) (2009 to 2016)	USA	[28]
N. T. Sagala & S. D. Permai	Enhanced Churn Prediction Model With Boosted Trees Algorithms in The Banking Sector	2021	XGBoost LightGBM CatBoost	88.8% - 91.3% 91.3% - 91.4% 91.3%	Bank (10000 customers)	Indonesia	[29]
O. F. Seymen & O. Dogan & A. Hiziroglu	Customer Churn Prediction Using Deep Learning	2021	Deep Learning Logistic Regression Neural Network	90.8% 89.3% 89.7%	A supermarket in Turkey (10000 customers)	Turkey	[30]
B. N. Bristy	Customer Churn Analysis and Prediction	2022	Hyper Parameter Random Forest Decision Tree K-Nearest Neighbor Logistic Regression	83.32% 85% 79% 79.5% 75.2%	Bank clients in Spain, France and Germany (10000 customers)	Bangladesh	[31]
S. Kim & H. Lee	Customer Churn Prediction in Influencer Commerce: An Application of Decision Trees	2022	Decision Trees	90%	Influencer marketing agency in Korea (100213 customers)	Korea	[32]
K. Matuszelan´ski & K. Kopczewska.	Customer Churn in Retail E-commerce Business: Spatial and Machine Learning Approach	2022	XGBoost Logistic Regression	62.5-65% 54.6-58.6%	Brazilian e-commerce site Olist (100000 orders) (2016–2018).	Poland	[33]