International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v8i1.2085
***Article Received**: 24 December 2021 **Revised**: 16 February 2022 **Accepted**: 22 February 2022 **Publication**: 31 March 2022*

Research-Based on Telecommunication in Mobile Service Provider's Performance using Enhanced Naive Bayes Classifier

Prof. Dharmesh Dhabliya
Professor, Computer Engineering Department,
Vishwakarma Institute of Information Technology,
Pune, India
dharmesh.dhabliya@viit.ac.in

Abstract
In recent years, mobile service providers have rapidly expanded across all countries. Considering unpredictable development trends, mobile service providers are essential to knowledge-based service businesses. Performance may be improved by creating and disseminating new information through innovation activities based on the usage of business intelligence. This research examined the performance of mobile service providers across all countries utilizing an enhanced Naive Bayes classifier based on telecommunication. In comparison to quantitative variables, the naive Bayes performs quite well. In the beginning, data is collected and the normalization technique is used for data preprocessing. Feature extraction is carried out using "Term Frequency and Inverse Document Frequency (TF-IDF)". "Decision Tree algorithm" is used for data analysis. Then the feature is selected using a two-stage Markov blanket algorithm. Enhanced Naïve Bayes Classifier is the proposed algorithm for telecommunication analysis and at last, the performance of the system is analyzed. This proposed algorithm is used to compare the mobile service provider's performances with existing algorithms. The proposed method measures the following metrics as Throughput, Packet loss, Packet duplication, and User quality of experience. The proposed algorithm is more effective and produces better results.

Keywords: *Mobile Service providers, Telecommunication, Enhanced Naive Bayes classifier, Term Frequency-Inverse Document Frequency (TF-IDF). Decision Tree algorithm, two-stage Markov Blanket algorithm*

I. INTRODUCTION

The growth of telecommunication has evolved in a change from a wired to a wireless network due to the fast development of telecommunication in the form of electronic networks, the internet, the phone network, etc. Telecommunication networks play a crucial role in linking the globe, and they can store and transmit a vast amount of information, both sensitive and non-sensitive, in the form of speech and text [1]. The telecommunication industry is now the fastest growing telecom market worldwide. The number of customers and the frequency with which they use mobile telecom services have changed dramatically as a result of more affordable services, increased penetration, a supportive government, and the continual decline in the level of the market. Communication technologies of the present day permeate every aspect of human existence. In addition, the widespread adoption of mobile phones in recent years is a result of the ease with which they may be used and the cheap cost at which they allow individuals of various socioeconomic backgrounds to stay in touch. There is a statistically significant correlation between a consumer's attitude about smartphone use and their gender, profession, degree of education, and household income [2]. A lot of pressure is being put on the telecommunications industry to improve and upgrade its service for the sake of its customers. The telecommunications sector has carved itself into a special niche. Consumers of mobile services are reportedly unhappy than those in any other sector. The worldwide market for internet, data, and broadband and phone services is expanding rapidly, and telecommunications providers are adapting and converging their network services to meet this demand. There are between 17 and 20 million annual users who migrate to a new mobile service provider. Because of the high levels of competition in the telecom industry, client loyalty is poor. Additionally, clients will migrate to a different service provider if they are not provided with a service that is both competitive and quality [3]. Mobile service providers have increased globally in recent years. Mobile service providers are vital to knowledge-based service firms given uncertain development tendencies. Using business intelligence to create and distribute fresh information may increase

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v8i1.2085
***Article Received**: 24 December 2021 **Revised**: 16 February 2022 **Accepted**: 22 February 2022 **Publication**: 31 March 2022*

performance. Hence, we proposed the enhanced naïve bayes classifier for telecommunication in mobile service providers.

Contributions of the paper

- 3333 records from the telecommunication churn dataset were gathered for the dataset used in this research.
- Normalization improves lead generation and data preparation quality used for data preprocessing.
- TF-IDF analyzes are also used in term balancing for feature extraction.
- Decision Tree algorithm" is used for data analysis
- Two-stage Markov blanket approach evaluates the statistics bounds of a process for feature selection.
- ENBC is the proposed algorithm for the performance improvement of the telecommunication analysis.

The remainder of the paper is structured as follows: In part II, a literature survey is provided. Described in Part III is a proposed approach. Part IV includes a result and discussion. Part V is the conclusion section.

II. LITERATURE SURVEY

This paper reviews several research papers and technical reports authored by diverse writers. Customers' propensity to transfer service providers within the telecommunications market is examined in study [4]. Data research revealed that customer switching behavior in the telecommunications sector is positively influenced by both words of mouth and core service failure. Research [5] used Porter's Five Forces Analysis to provide the first complete analysis of the competitiveness of Kenya's telecommunications sector. It is one of the few that has been presented and has the potential to offer useful information and suggestions for the leaders of Kenya's mobile service providers as they work to create sustainable competitive and survival strategies. Study [6] analyzed the customers' attitudes about mobile money affect their propensity to return customer continuance intention (CCI). They were put to the test using PLS-SEM, and a suitable sample size of 507 mobile money users was used for the test. CCI was found to be affected by factors like user satisfaction, user trust, and mobile money use. To the state of mobile service quality in Italy, the specific, the Delphi method is first considered to finalize a quality management structure of mobile telecommunication services (MTSs) features, indicators, and drivers, as described by the study [7] based on thorough fundamental sources of information for the field, and to select the critical components with comparison to the Italian context. The goal of study [8] was more about the mobile telecommunications industry's technology life cycle and improves the accuracy of future predictions. Specifically, data was collected in Turkey to add to the body of knowledge in engineering management. Author [9] provided the factors that contribute to a positive experience with a telecommunications provider are the focus of the current data. To calculate an overall customer experience management index (CEMI) from observations, they apply a genetic algorithm (GA) based approach for weighting distinct qualities of service. The research [10] used the SERVQUAL model to evaluate the quality of service provided by two major telecom providers in the Sultanate of Oman, Omantel, and Ooredoo, and to examine the impact of five SERVQUAL characteristics on the loyalty attitudes of those who use their services.

III. PROPOSED WORK

This research is based on telecommunication in mobile service providers' performance using an enhanced Naive Bayes classifier. A mobile service provider (MSP) is a business that offers consumers wireless internet and voice connections for their mobile devices. Telephone and comparable services have historically been supplied by a telecommunication company. Companies that provide mobile wireless communication services are included here with dominant local exchange carriers as well as competitive different exchange providers. Figure 1 depicts the Schematic representation of the proposed methodology.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v8i1.2085
***Article Received**: 24 December 2021 **Revised**: 16 February 2022 **Accepted**: 22 February 2022 **Publication**: 31 March 2022*

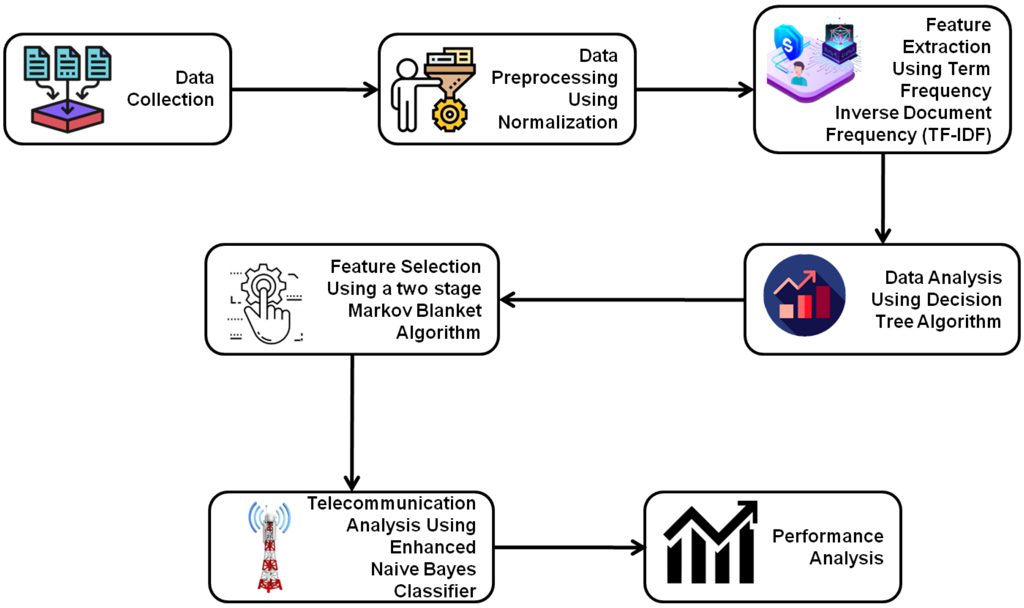

Figure 1: Schematic representation of the proposed methodology

A. Data collection

The Kaggle sample online dataset provided the telecommunication churn dataset that was utilized in this study. There are 21 variables and 3333 records in the dataset. There are 3333 records in all, and 21 different categories to categorize them into. The telecommunications churn market shows that there are twenty determinants and a single outcome in the dataset (churn). There are much fewer churners than non-churners in the sample, which is very unbalanced. There are 3333 total occurrences, and only 483 (14.5%) of them are churning.

B. Data preprocessing using normalization

Because the device data set has some inaccurate or missing information, as well as some overlap in form during individual potential by both data sets, we do data-preprocessing, which includes normalization and dealing with all such issues. The complexity of telecommunications was reduced by using an equation to normalize data for the mobile service GH^{nq} inside the domain [0, 1]:

$$Data\_X_{norm} = \frac{GH^{nq} - Data\_X_{min}}{Data\_X_{max} - Data\_X_{min}} \times [max_{value} - min_{value}] + min_{value} \quad (1)$$

where $Data\_X_{norm}$ stands for the normalized value of a dataset, $Data\_X_{min}$ stands for the minimum value of a dataset, $Data\_X_{max}$ stands for the maximum value of a dataset, GH^{nq} stands for the original value of the mobile service data source, max-value and min-value stand for the range of a normalized input data, with max-value = 1 and min-value = 0, and *GH* stands for the initial telecommunication in mobile service data.

C. Feature extracting using term frequency-inverse document frequency

TF-IDF is the oldest and most traditional. It's built on telecommunication in mobile service, where a document is represented by a list of its terms. Similarly, TF–IDF posits that if a term is crucial to a text, it will occur often inside that document while appearing seldom elsewhere. Assumption (a) is connected to the TF, whereas (b) is linked to the IDF. Term *k* significance is measured by its frequency of appearance in document *l* (denoted by the parameter me_{ji}), with higher values indicating greater frequency. The parameter dfi is the total number of papers that include the term i at least once; the higher the value, the more often the term occurs in mobile service. For term i to be deemed significant in document j, it must have a high TF (me_{ji}) and a low DF (DF_k). Therefore, the TF–IDF is defined as follows:

$$\mathrm{TF-IDF}_{kl} = \mathrm{me}_{ji} \times \log\left(\frac{\mathrm{R}}{\mathrm{ce}_j+1}\right) \quad (1)$$

Where $\mathrm{IDF}_k = \log\left(\frac{\mathrm{R}}{\mathrm{ce}_j+1}\right)$ is the IDF's smoothed version of the document frequency of term k.

And R is the total number of documents in the collection, and me_{ji} is the frequency of term k in documents. For the

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v8i1.2085
***Article Received**: 24 December 2021 **Revised**: 16 February 2022 **Accepted**: 22 February 2022 **Publication**: 31 March 2022*

dataset, we select terms that have, on average, TF-IDF score values.

D. Data analysis using decision tree algorithm

Decision tree is a method for grouping and classifying views. This technique finds the best way to group instances. Gradually isolating applicant categories until none are left creates a numeric difference. If the decision tree is too big, trim branches. Strategic splits affect the tree's dependability. Decision tree algorithms have unique criteria. A decision tree is created when an algorithm chooses how to split a node. Sub-nodes are becoming more similar. The purity of the node is assumed to increase as the dependent variable does. The computed tree selects the most stable sub-node division by dividing each node into as many sub-nodes as possible given all relevant criteria. The algorithm 1 depicts the decision tree.

Algorithm 1: Decision tree

Design tree

Launch root node data

Select an attribute and construct a logical attribute check

Transmit a subset of samples that pass the test and fulfill each result to the child node that is associated with it. •

Child nodes should be retraced in their entirety

And so on, until the leaves are either "clean," with just one example of a given class, or "nearly pure," with the vast majority of the samples being of the same class;

Plum tree

Eliminate nodes if they aren't contributing to better classification precision.

Don't over fit, meaning don't use a training set full of identical items.

The universe isn't linear. Data show tree models map nonlinearity successfully. Decision trees are popular big data techniques. Decision tree learning is supervised. Decision tree algorithms may be utilized for more than only training data. Implementing a Choice Tree creates a training model that can forecast target variable class or value based on decision data rules. Choosing the initial node and subsequent levels of a decision tree is the hardest task. Decision trees reflect human brain processes to make complicated data sets actionable.

E. Feature selection using two stage Markov blanket algorithm

Two-phase feature selection and Markov Blanket Discovery are used. Two-stage feature selection split the procedure into two steps. First, experts' suggestions were utilized to decrease characteristics, and then Markov blanket discovery was applied. In the Markov blanket, the "churn" target attribute T was established with a blanket including reliant qualities; hence, MB (T) is a minimal collection of attributes. Using IAMB, non-zero Markov blanket members were found. We've defined criteria for differentiating dependent and independent attributes. The Markov process finished when no new characteristics could be introduced. IAMB reduces features one by one using a non-traditional MB (T) method. The two-stage feature selection method outperforms expert analysis. Each mobile service database has its numbers and characteristics, thus it won't work with all of them. In this study, we categorized turnover detecting skills. The discriminating power of the independent variables was measured using AUC. If the ROC curve is slanted to the lower left, the model is more sensitive and has greater classification accuracy. This study uses AUC to select features. When the variable's AUC was larger than 0.5, its number was computed. The ideal number of predictors was determined by minimizing mutual information between variables. If the model's performance is sufficient after this threshold, feature selection is complete; otherwise, further thresholds must be applied. AUC formula:

$$V = \frac{\sum_{1=0}^{p-1}\sum_{i=0}^{r-1}(a_j,b_i)}{pr} \quad (2)$$

$$J(a_j, b_i) = \begin{bmatrix} 1\ if\ (a_j > b_i) \\ 0\ otherwise \end{bmatrix} \quad (3)$$

Using mutual information, variables were compared. Variables with high mutual information exchange a lot of data. This variable had no corresponding data, hence it was removed. This data exchange determined the two variables' connection. Using mutual information, variables with AUCs < 0.5 were eliminated. This approach was shown to be very useful and effective for feature selection, but it was insufficient for datasets. Similarly, a two-stage feature selection approach was presented. Following are the procedures the source used to filter the characteristics in two stages.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v8i1.2085
***Article Received**: 24 December 2021 **Revised**: 16 February 2022 **Accepted**: 22 February 2022 **Publication**: 31 March 2022*

Step 1: The first step is to type in the real characteristics and see whether they fall within the structural model or the text model category.

Step 2: The second, if the functionality is part of the structural module, implement it. Use the evaluation function to discard features that don't pass. The feature assessment function scores the text model's features. Next, high-scoring features are picked. This level adds a feature subset.

Step 3: The third, the wrapped module applies learning techniques to a subset of features. Final feature selection enhanced parameters using measuring criteria. Two-stage feature selection is better than one.

F. Telecommunication analysis using enhanced Naïve Bayesian classifier (ENBC)

Data classification uses the ENBC approach. ENBC uses Bayes' theorem to classify data. Like decision trees and classification approaches, this strategy offers good prediction performance. The telecommunication analysis data is used to train the ENBC, a Bayesian approach. To be more specific, ENBC aims to maximize the posterior probability of the class variable given attributes.

$$\underset{d\epsilon D}{\operatorname{argmax}} M(D|b_1, \ldots., b_r) \qquad (4)$$

In classification, the Bayes rule is expressed as:

$$M(D|b_1, \ldots., b_r) = \frac{M(D)M(b_1, \ldots.., r|D)}{M(b_1, \ldots.., b_r)} \qquad (5)$$

As the ENBC presumes that each attribute is conditionally independent given the class variable, we must reformulate Eq. (6) as follows. Figure 2 shows the structure of naive Bayes.

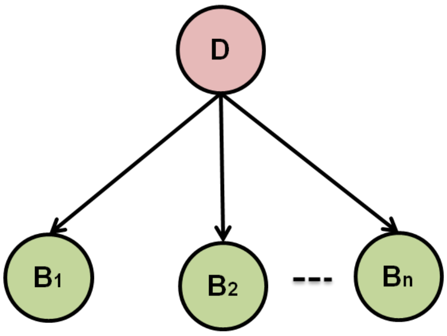

Figure 2: Naive Bayes structure

$$M(D|b_1, \ldots., b_r) = \frac{\prod_{j=1}^{r} M(b_j|D) M(D)}{\sum_d \prod_{j=1}^{r} M(b_j|D = d) M(D=d)} \qquad (6)$$

No matter what the value of the class variable is, the denominator remains the same. As a result, equation (7) may be reduced to:

$$\underset{d\epsilon D}{\operatorname{argmax}} M(D) \prod_{j=1}^{r} M(b_j|D) \qquad (7)$$

IV. RESULT AND DISCUSSION

In this paper, we proposed the enhance naïve Bayesian classifier for telecommunication in mobile service providers. Experiment used throughput, packet loss, packet duplication, and user quality of experience. Existing methods such as k-nearest neighbor [KNN], Cart algorithm [CA], support vector machine [SVM], and multi-objective rain optimization algorithm [MOROA] are compared to the proposed work [ENBC].

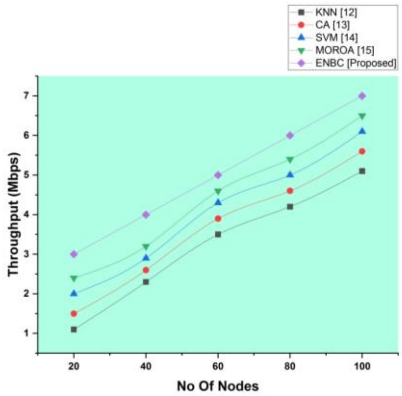

Figure 3: Comparison of the throughput (Mbps)

Figure 3 depicts the number of nodes on throughput. In data transmission, throughput is the quantity of data transported effectively in a certain period, measured in megabits per second (Mbps). With higher data transmission levels, a network's performance is improved. The proposed work has the greatest 7 Mbps of throughput than that of the existing approaches throughout this examination.

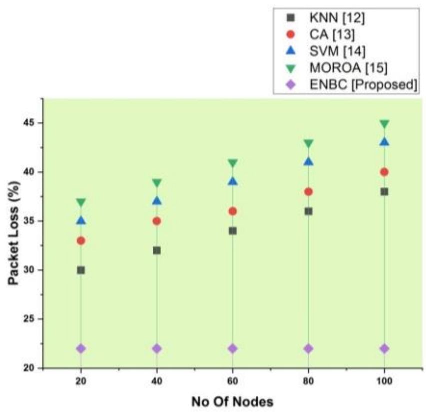

Figure 4: Comparison of the packet loss

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 8 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v8i1.2085
***Article Received**: 24 December 2021 **Revised**: 16 February 2022 **Accepted**: 22 February 2022 **Publication**: 31 March 2022*

Figure 4 depicts the number of nodes on packet loss. A system may experience packet loss if any of the data being transferred is corrupted or otherwise lost before reaching its intended recipient. Missing packets during data transmission may be caused by a variety of circumstances, including but not limited to network problems, hardware problems, and software defects. The proposed work has a minimal 22% of packet loss than the existing method like KNN (38%), CA (40%), SVM (43%), and MOROA (45%).

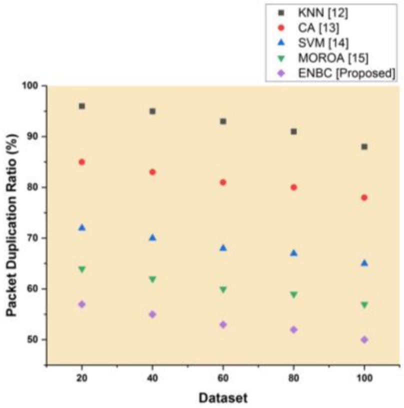

Figure 5: Comparison of the packet duplication ratio

Figure 5 depicts the comparison of the packet duplication ratio. The network packets that are being passed to analytical tools for monitoring, analysis and recording are reduplicated using packet duplication technology. We evaluate the suggested work [96% of ENBC] in comparison to state-of-the-art techniques like KNN (88%), CA (78%), SVM (65%), and MOROA (57%). Throughout the testing, it showed the lowest packet duplication compared to the other available methods.

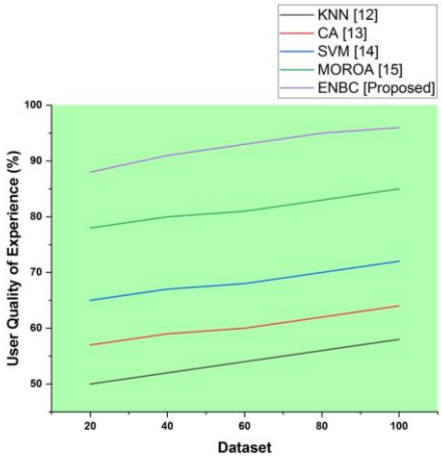

Figure 6: Comparison of the user quality of experience

Figure 6 depicts the comparison of the user quality of experiences. The term "quality of experience" (QoE) refers to how satisfied or frustrated a consumer is with a service. Originally developed in the telecommunications industry, user QoE takes a broad view of the customer's interaction with a service. Throughout this investigation, the proposed work ENBC (96%) has highest user quality of experiences than the existing methods like KNN (58%), CA (64%), SVM (72%), and MOROA (85%).

Discussion

The figures show how the suggested technique is an improvement over the current system, which has several general drawbacks. The following are some of the drawbacks of the methods now in use. The proposed method is compared to the existing methods like k-nearest neighbor [KNN], Cart algorithm [CA], support vector machine [SVM] and multi-objective rain optimization algorithms [MOROA] are more efficient for the ENBC for telecommunication in mobile service providers. As KNN the number of variables raises so does the size of the input dataset. Finding relationships to all the neighbors in a larger data dataset and then determining which one is closest consumes a lot of effort. At the moment CA, telecommunications firms are unable to detect churn which is the outcome of industry action. The SVM for reducing class imbalance in the original dataset, businesses may aggressively remove them from their user lists and manage the risks associated with this kind of dropout. MOROA customer retention is conducted to prevent customer problems, a key issue in telecommunication sectors. User expectations increase, competition between businesses in the telecommunication industry increases, and organizations implement a wide variety of processes and methods. The experiment used that metric, along with throughput, user experience quality, packet loss, and packet duplication. As a result, the proposed work for the parameter performs more effectively in terms of throughput and user experience quality, while mobile service providers perform least effectively in terms of packet loss and packet duplication ratio.

V. CONCLUSION

This study's contribution is that the telecommunications sector may place more attention on improving expression and network service breakdown in order to obtain an edge in highly competitive industries. The paper proposed the ENBC for telecommunication in mobile service providers. The research analyzed 3333 records collected from the telecommunications churn market. Normalization was used for data preprocessing, Feature extraction was carried out using “Term Frequency-Inverse Document Frequency (TF-

IDF)", a Decision Tree algorithm was used for data analysis, and two stages Markov blanket algorithm was used for feature analysis. The experimental results are provided as 7Mbps of throughput, 22% of packet loss, 50% of packet duplication, and 96% of user quality experiences. This research has certain limitations, the most significant of which is its limited focus on the factors that may have previously influenced consumers' propensity to switch brands. Additional variables that may influence customer switching behaviour should be included in future studies. It's possible that any mobile service provider, due to the growing telecommunications industry, may dominate in the next few years in terms of what matters most to customers.

REFERENCES

[1]. Says, S., 2019. DDOS attack detection in a telecommunication network using machine learning. Journal of Ubiquitous Computing and Communication Technologies (UCCT), 1(01), pp.33-44.

[2]. Nekmahmud, M. and Rahman, S., 2018. Measuring the competitiveness factors in telecommunication markets. In Competitiveness in emerging markets (pp. 339-372). Springer, Cham.

[3]. Bhatti, H.S., Abareshi, A. and Pittayachawan, S., 2019. Factors that Impact Customers' Loyalty for Mobile Telecommunication Products and Services in Australia. In CONF-IRM (p. 9).

[4]. Lim, K.B., Yeo, S.F., Goh, M.L. and Koh, W.M., 2018. A study on consumer switching behavior in the Telecommunication industry. Journal of Fundamental and Applied Sciences, 10(6), pp.1143-1153.

[5]. Chesula, O.W. and Kiriinya, S.N., 2018. Competitiveness in the telecommunication sector in Kenya using Porter's five forces model. International Journal of Research in Finance and Marketing (IJRFM), 8(7), pp.1-10.

[6]. Yeboah-Asiamah, E., Narteh, B. and Mahmoud, M.A., 2018. Preventing customer churn in the mobile telecommunication industry: Is mobile money usage the missing link?. Journal of African Business, 19(2), pp.174-194.

[7]. Lupo, T. and Delbari, S.A., 2018. A knowledge-based exploratory framework to study the quality of Italian mobile telecommunication services. Telecommunication Systems, 68(1), pp.129-144.

[8]. Kalem, G., Vayvay, O., Sennaroglu, B. and Tozan, H., 2021. Technology forecasting in the mobile telecommunication industry: A case study towards the 5G era. Engineering Management Journal, 33(1), pp.15-29.

[9]. Khan, N., Akram, M.U., Shah, A. and Khan, S.A., 2018, May. Calculating customer experience management index for telecommunication service using genetic algorithm-based weighted attributes. In 2018 IEEE International Conference on Innovative Research and Development (ICIRD) (pp. 1-8). IEEE.

[10]. Belwal, R. and Amireh, M., 2018. Service quality and attitudinal loyalty: Consumers' perception of two major telecommunication companies in Oman. Arab economic and business journal, 13(2), pp.197-208.

[11]. Li, K.G. and Marikannan, B.P., 2020. Hyperparameters Tuning and Model Comparison for Telecommunication Customer Churn Predictive Models. In 3rd Global Conference on Computing & Media Technology (pp. 475-83).

[12]. Wibawa, D.W., Nasrun, M. and Setianingsih, C., 2018, December. Sentiment Analysis on User Satisfaction Level of Cellular Data Service Using the K-Nearest Neighbor (K-NN) Algorithm. In 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC) (pp. 235-240). IEEE.

[13]. Rai, S., Khandelwal, N. and Boghey, R., 2020. Analysis of customer churns prediction in the telecom sector using cart algorithm. In First International Conference on Sustainable Technologies for Computational Intelligence (pp. 457-466). Springer, Singapore.

[14]. Chory, R.N., Nasrun, M. and Setianingsih, C., 2018, November. Sentiment analysis on the user satisfaction level of mobile data services using the Support Vector Machine (SVM) algorithm. In 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS) (pp. 194-200). IEEE.

[15]. Pustokhina, I.V., Pustokhin, D.A., Nguyen, P.T., Elhoseny, M. and Shankar, K., 2021. Multi-objective rain optimization algorithm with WELM model for customer churn prediction in the telecommunication sector. Complex & Intelligent Systems, pp.1-13.