International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 6, June 2020, pp. 534-544, Article ID: IJARET_11_06_049
Available online athttp://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=6
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.11.6.2020.049

SVM BASED CHURN ANALYSIS FOR TELECOMMUNICATION

R. Sudharsan
Research Scholar, Department of Computer Science & Engineering,
Vels Institute of Science, Technology & Advanced Studies,
Chennai, India.

Dr. E.N. Ganesh
Dean, School of Engineering,
Vels Institute of Science, Technology & Advanced Studies, Chennai, India.

ABSTRACT

Data mining is a technical analysis for the massive database with meaning full output. Churn analysis is best applications for the data mining able to determine the behavior of the customer in changing different services. The available marketing tool has the limitation to predict the changing behavior for the customer. This paper discuss about the churn analysis in telecommunication sector for the 2019 year Q2 period. Machine Learning is an advanced development in data mining to extract the features from large quantities of data. The paper discuss about supervised machine learning model. The supervised model designed by support vector machine (SVM) classification steps for two group separation of churn customer and non churn customer. The model aim to analysis total number of subscriber in voice call routing along the period. The churn analysis can predict the churn rate and the probability for month and the year. The proposed method, classify the customer from churn and non-churn by SVM increase in accuracy for the existing system. The proposed work implemented using MALAB R 2014b simulation software and the results were discussed.

Key words: Telecommunication, churn analysis, data mining, subscriber call, support vector machine.

Cite this Article: R. Sudharsan and Dr. E.N. Ganesh, SVM Based Churn Analysis for Telecommunication, *International Journal of Advanced Research in Engineering and Technology*, 11(6), 2020, pp. 534-544.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=6

1. INTRODUCTION

Nowadays customer thinks about the long distance communication in related to products and services. Application of telecommunication services increases day by day. The telecommunication services facilitate the customer in the economy and entertainment wise.

The choice of customer over the services offers the services package for the customer preference level. The customers spread in all parts of the country. They are connected over the networking technology. The telecommunication served over the networking equipments and devices. The range of communication distance covers the different network service equipment. The telecommunication services are of two types namely wide band and narrow band services. Application related with the telecommunication like telephony, video conferencing and facsimile, broad cast, television, web application like Internet oriented communication and data transmission. Increasing the telecommunication applications the reliability maintains using the monitoring of the network equipments. The complexity rises in maintain the network performance among the customer. The network architecture increases the complexity for the customer choice in selecting the branded network. The telecommunication products and package has the competition in maintaining the customer choice. So the communication services offer the various limited package to withstand the customer choice [1].

The technology development changes from the generation by generation. The mobile service gets transformed from the Landline, cellular and mobile network with 2G, 3G, 5G with Voice oriented Long Term Evolution (LTE). The mobile phones calls are routed and forward through Virgin Media (BT) services. The incoming and outgoing calls are travel through the BT (Base Station) services. The internet access, mobile access shared over the BT services. It maintains all the services records among the network [2]. Network equipment provider companies take care about the services status about the end to end services, which able to predict the customer life time value (LTV). The LTV gives the information about the churn prediction for the company. This approach uses the concept of micro segmentation, macro segmentation for the customer among the services. The accuracy for the churn prediction methods has the limitations in extending the network. The churn prediction over the mobile network is the challenging task. The paper focus on the mobile operator trying to retain the customer and try to satisfy their needs [3]. The churn analysis estimated over the data period still the customers are present in the company. The company focuses to retaining the customer. The company in the needs to identify the churn customer, retaining customer , strategies for the profitable customer and customer retention over the telecommunication over the fixed and the wireless based communication like broad band, voice and the services [4]. The extension of customer increases the need for the company to rise of technology services like spectrum allocation for the network. The telecom services like fixed (landline) and wireless services like satellite communication. The long range services are access through satellite communication. The company increases the technical facilities to with stand the customer services and the extension of services [5].

The paper focuses on the drawbacks of churn prediction and the existing human error and statistical method. The proposed method overcomes the customer data misclassification, missing value etc. The data misclassification, missing value, customer information in massive data base overcome by the feature extraction process, which specifically catches the customer information. This feature reduces the high dimensional data into the workable data format. The misclassification of churn and non churn customer accuracy increases using the churn analysis. The proposed work uses the massive data like telecommunication the period of Q2 for the year 2019 collected from the telecommunication dataset.

A company under pressure to retain the customer over various offers likes advantages and services. The author Michael C. Mozer et al., 2010 [6] proposed technique related with the machine learning. The technique predicted the churn rate and determines the subscriber retention by the factor like incentives, money value packages and maintaining the maximum success for the carrier. The factor influencing the profitability for the carrier willing to spend in retaining the customer. The influencing factor for the churn prediction reduced by the proposed author Emiliano G. Castro at al., 2015[7] focused on the regularity examination approaches based on the feature extraction from the logical records for prediction of the churn modeling. The algorithm uses the data conversion format here the real time data conversion for the length of data in fixed range of array with the different models. The input is trained using the classifier techniques like probabilistic, K nearest neighbour and other machine learning techniques. The performance of the classifier predicted using the performance metrics. This method focused on data obtained from the electronic commerce and services related with the on line. The author Ning Lu et al., 2014[8] implemented customer churn model enhances the customer churn rate. The logistic regression method researches the two common factors like the starting learners and various churn predicting methods to build the cluster respectively. The customer expectation not satisfied by the company. The author Bingaman Huang et al., 2012 [9] describes a customer experience request. A size of request list be huge must be difficult to handle the service orders. The details related with these three services order like rental charges due date, approving services with new feature. This information relates with the customers of other countries relates with the various features. Marcin Bienkowski et al.,2018[10] presented the problem in the temporal activity for accounting the number of information systems FIB indicates the Forwarding Information Based update the router for CPU load low and reduced updates transferred among the software attached to network build with the controller act as the remote switch.

A Significant Market Power (SMP) is the level of failure or dominance market analysis for the UK country market. A market condition subjects to the business narrowband and broad band services. Another rule relates with the fixed telephony and the market relate with the mobile phone call. The technological factor improves the company in understanding investments plan. The investment plans has the high level for the customer over the churn in the business. The research take part in differentiating the customer in identifying the churn other services providers. The unregulated telecom companies may regulate the competitors in the market and the customer has the choice in their needs. The drawbacks for the existing method in the customer data dimension reduction method. The feature selection becomes problem in detecting the specific feature. The problems in data classification process. The proposed method reduces the data dimensionality and increases the churn prediction accuracy. The method reduces the complexity in the training data.

2. MATERIALS & EXPERIMENTAL PROCEDURES

2.1. Materials for Churn Analysis for Telecom Market Data

The percentage of churn prediction for the sub for the subscribers in a service may discontinue the service for the period of time. A company may able to large coverage of their client with its growth rate. The number of new customer serves exceeds the current churn rate, which considers the telephone and mobile phone service industry. The area location for the several companies was connecting the customer. The company makes it easy for the people in transfer from one service to other services industry. The increasing pressure for the competitors and the centralized mandates for improving the retention rate for the customer profitability, which becomes increases the retention rate for the profitable customer. The action increases urgently to service providers for telecom sector [11].

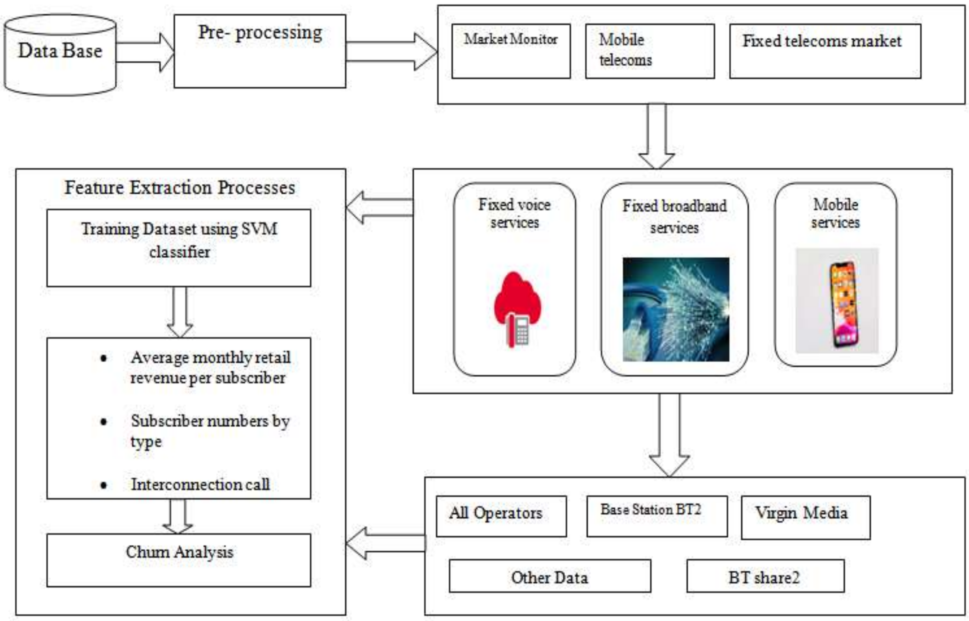

Figure 1 Churn analysis for Mobile services in Telecom Market.

The figure 1 represents the proposed model for the analysis of churn in the services of mobile services with telecom market. The method involves the action for the database high dimensionality reduction, the pre- processing for the telecom services. The telecom services for fixed voice services, broad band services and mobile services. The high dimensionality data reduced to the three sections namely market monitor, mobile telecom and the fixed telecomm market. The feature extraction process involves the training dataset using support vector machine. The data sector analysis involves the factor likes the all operator, base station, virgin media, other data and BT shares.

The below services are taken for the analysis

a. Fixed voice services

The UK country has the fixed voice service with the revenue £1.8bn for the year Q2 2019. The value decrease near to 1.8% from the previous quarter. The 6.2% for the Q2 2018 year. The Base Terminal (BT) shares for this revenue with 41% down by the 2.5 percentage point's year wise.

b. Fixed Broad Band services

The UK fixed broad band connection at the end of Q2 2019. The percentage previous from 1.5% from Q2 2018.

c. Mobile services

The telephone services (mobile) generate the revenue in Q2 2019 decreases by 0.98% from the previous year.

2.2. Methods

2.2.1. Database Collection

The telecommunication services providers has to increase their IT services. The connection infrastructure concentrates on the provision data and services for voice with the high quality, reliable and affordable. The security over the network becomes majority priority in the telecom services. The security for the network becomes the major priority in the telecommunication sector. The challenges in the tragedy are encourage by the novel technologies [12]. Some telecommunication market data segments reaches up to 6.5% increased by the 2014. The mobile telecom market data analysis for the mobile network with the voice, innovation and data along the network services provider broadband and the video operator and serviced managed [13].

2.2.2. Pre-processing

The data mining techniques for the data pre-processing involves the transformation of raw data into machine learning format. The raw data is the inconsistent, lacking with the behavior or trends with many error. The data processing method involves the certain issues like data noise removal, feature selection, standardization, transformation etc. The training for the final stage gives the product for the data incomplete, the lacking with attribute values of interest or containing the accurate data using table 1 and table 2.

Table 1 Revenue report by Mobile Telephony.

Consolidate by Year	Data for access services	Data for Fixed call service	Data for Online mobile calls	Data for Off line - mobile calls
2017	16,168	12,147	253	242
2018	13,834	10,193	208	224
2018 Q2	3,444	2,545	53	55
2018 Q3	3,517	2,588	54	58
2018 Q4	3,478	2,574	50	55
2019 Q1	3,392	2,565	44	51
2019 Q2	3,411	2,618	43	52

Table 2 Revenue report by network access.

Consolidate for the year	Data for Internatio nal calls	Data for other network calls	Data over Short and multimedia message	Data other services
2017	296	416	436	645
2018	243	377	343	610
2018 Q2	58	93	87	150
2018 Q3	59	102	93	154
2018 Q4	64	93	75	157
2019 Q1	57	81	65	148
2019 Q2	54	72	62	142

The noise data contains the error or outer linear.

The algorithm steps in data pre-processing

Step 1: The library functions to be import

Step 2: Data set to load

Step 3: Normalized the value for missing.

Step 4: Categorize the missing value.

Step 5: The dataset can be sub divided into the training and testing data.

Step 6: A feature extraction process involves the Scaling function.

Step 7: Predicting the churn and non-churn customer minutes/ message/PB.

Average value for the monthly retailer revenue for individual subscriber. The table 1 and 2 indicates the estimated detail using the revenue profit over the telephone mobile. The data representations presented over the access and bundled call for the services. The table 2represents the international call and other call relates with the data from the short message service and multimedia message services. The table data year 2018-2019 the churn prediction in month, year and average monthly revenue for individual subscriber.

2.2.3. Support Vector Machine Classifier (SVMs)

It is the supervised learning methods used for the two level of process like classification, regression task, which originated through the statistical learning theory model. The classification steps involve the feature vector for global classification models. The data generates the overlapping, partitions based on non- overlapping generates all the attributes. The classification process becomes the activity like the training steps and the testing steps involve the f or target values (class labels) and the different attributes features. An objective for the classifier capable in extracting the objective values for the instance of data in the trying set and known attributes. A missing data able to classification related problem which can be viewed as the two- division of problem with the ones objective able to separate the two class functions [15]. The classifier able to generate conditions were the various linear classifier able to separate the data with the condition of maximizes the distance between the current data points and near data points termed as the optimal separating hyper plane. The classifier generates the better options, which is the basic concept for the classifier based on the SVM, which approaches to select the hyper plane for the maximum margin as shown in figure 2.

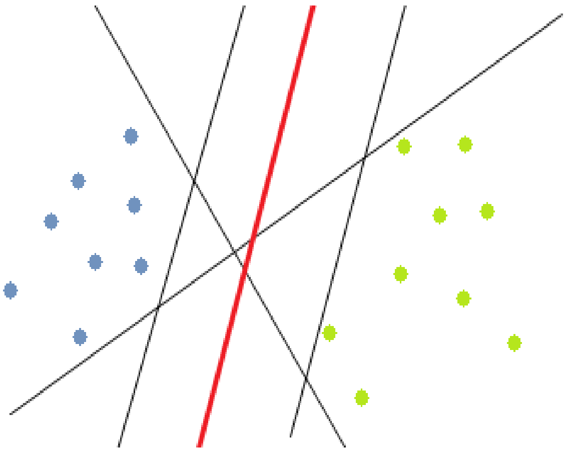

Figure 2 Hyper plane pattern for data separation.

The hyper plane deicide the data points belong to the group. The SVM relays on the pre-processing the customer information pattern in a high dimension with the original data feature space. The data categorized into two separate hyper planes with the linear approximation mapping with the sufficient huge data size reduction.

2.2.4. Calculation for the Churn Analysis

The churn rate is the attrition rate for the wide sense measures for the number of individual items collected over the group of specific period of time. The two most important factors can resolve the level of the steady state value for the customer in the business supports.

Churn Rate = (user at the beginning – User at the end) / Users at the beginning. (1)

The equation (1) termed as the business applied with the contractual customer base. A service model based on the subscriber telephone network and the TV operator. It may be termed as the active customer turn over in peak to peak network, simulator, which measures the go back the investments in marketing for the hybrid model related with marketing.

The churn rate for month = $\frac{User\ lost\ this\ month}{(user\ at\ start\ of\ month+Users\ added\ this\ month)}$ (2)

The churn rate describes the total count of customer transferred for a specific period of time. The churn rate for the 12 months period to the mean count of customer occupied over the month.

Churn rate for the year = $\frac{User\ Lost\ this\ Year}{(Uers\ at\ start\ of\ year+Users\ added\ this\ year)}$ (3)

The calculation for the probability churn for month basis might start from the count of users with the model churn for the month. The total number of users per day for the month, which able to get the number of churn per day. The divisions for the total number of users per days which are multiplied the counted days can results in the churn rate for month. A churn predicted data compressed using the Cohort analysis to reduce the churn numbers, which covers the specific reasons where the user gone and the action make withholding using the attainment control, action and time.

3. RESULTS AND DISCUSSION

A designed work is simulated with the help of MATLABR 2014b software. An input data set can be analysed and their description obtained over an extraction of feature and training over the dataset.

3.1. Description about Dataset

The dataset selection for the analysis for the telecommunication is Q2 2019. A count of active user's mobile subscriber for 85 million over the end period of the Q2 2019 for the year. The total outgoing voice call through mobile in minutes for 41 billion achieved through Q2 period down for the 0.3 billion (0.7 %) for the year.

3.2. Feature Extraction

The data collection for the preprocessing stage for the entire operator. A feature extraction data for different operator, Base terminal parameter. Data collected over the over the month of January to March as Q1. The month of April to June as Q2 and data Q3 represents the data from the July to September. An October to December month data as Q4.

3.3. Training Dataset

The SVM helps in analyzing the data need for the process of classification and analysis for the regression. The trained details for the 2018 period of March to December data, which shows the average number of customer over the period.

Figure 3 Monthly Average Revenue per subscriber.

The detail shows the customers related with the post pay, pre- pay and subscriber at edge of period. The net duration for changing the mobile subscriber over the time period.

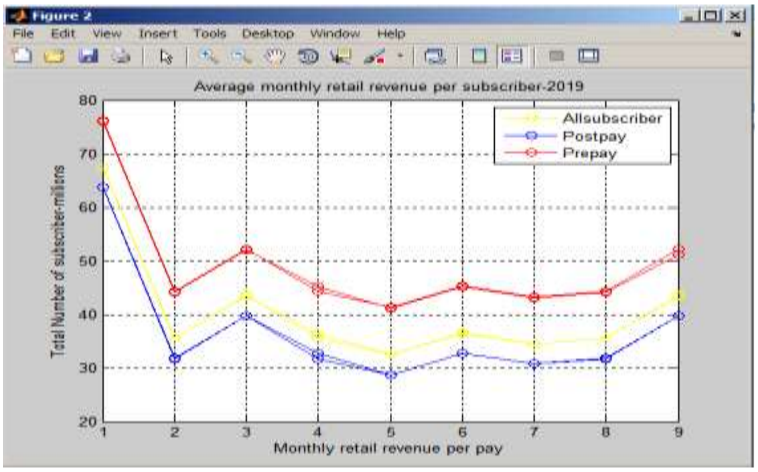

Figure 4 Monthly Retailer Revenue per subscriber for 2018.

The figure 3 and figure 4 indicates average retailer revenue for individual subscriber over the time period 2018-2019 shown over the figure. Average revenue for the year 2017 value is 26.00 and 2018 for the year is 27.00. The average value for the year results with millions of subscribers present for an analysis.

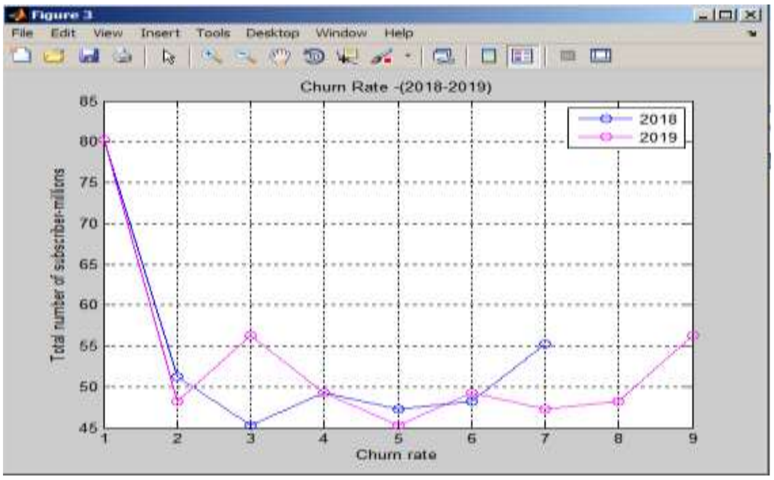

Figure 5 Churn Rate Analysis for 2018-2019.

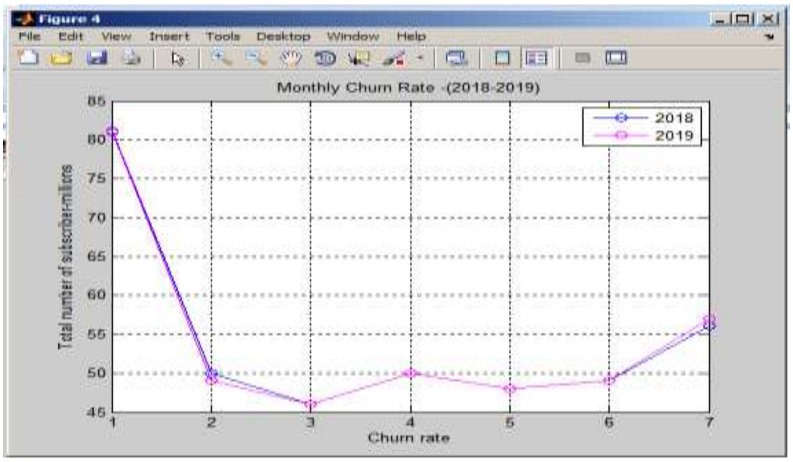

Figure 6 Monthwise Churn Rate 2018-2019.

The figure 5 and figure 6 indicates the monthly rate churn analysis for the year 2018-2019.A rate of attrition or churn customer stops in extending for the business for an entry. The churn rate for the year Q2 2018 value is 0.7255, 2018-Q2 is 0.7196, 2018 - Q4 is0.782. The figure 4 shows the month value is monthly rate with churn is 2018-Q2 is -0.2018-Q2 is -0.0288, 2018-Q3 is 0.0188, 2018-Q4 is -8.3963e-04 and 2019-Q1 is -0.0055.

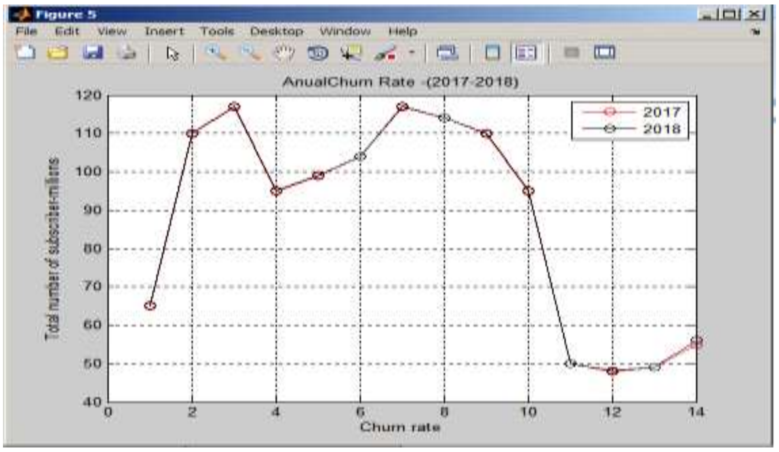

Figure 7 Annual Churn Rate 2018-2019.

An annual churn rate for the year 2017-2019, which might be able to determine the annual churn rate for 2017 is .0675 and 2018 is -0.0436.

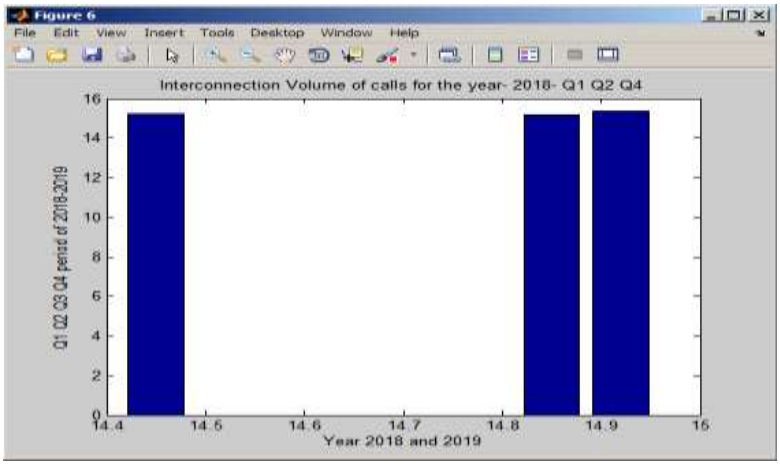

Figure 8 Interconnection calls volumes for the year 2018-2019.

A business needs the customer churn with as low as possible. The business needs to keep the customer within the company. The previous year's shows the loss of customer for the company. So the churn rate increases in the previous year. The annual churn rate makes the decision hardly for the maintaining the customer. The data for the year predicted over the correct churn plan. The figure 4 indicates the telephone access for the 2018-2019 year. The probability for the monthly churn for the year around 30%.

4. CONCLUSIONS

The designed work implies the churn analysis in terms of subscriber for the month and the year wise. A data collection based on the churn analysis for the telecommunication market for the update of Q2 2019. The services for mobile services describes the data presents the number of message passed through the mobile message with the SMS and MMS can continue over the period up to 32.0% for the year 2019. The monthly churn analysis and the annual churn analysis for the year 2017-2018 might be obtained. The future work is carried over the churn analysis for the fixed broad band and voice services. The evaluation for an association for the purchaser defeat rate to reduce it. The churn customer decreases with the facility for technical analysis increases.

REFERENCES

[1] Umman Tuğba Şimşek Gürsoy, (2010) Customer churn analysis in telecommunication sector, İstanbul Üniversitesi İşletme Fakültesi Dergisi Istanbul University Journal of the School of Business Administration, 39, pp.35-49.

[2] Kristof Coussement, Dirk Van den Poel, (2006) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, pp.55. http://wps-feb.ugent.be/Papers/wp_06_412.pdf

[3] Eunjo Lee, Boram Kim, Sungwook Kang, Byungsoo Kang, Yoonjae Jang, Huy Kang Kim, (2018) Profit Optimizing Churn Prediction for Long-term Loyal Customer in Online games, IEEE Transactions on Games, 12(1), pp.41-53.

[4] Pretam Jayaswal, Bakshi Rohit Prasad, Divya Tomar, and Sonali Agarwal, (2016) An Ensemble Approach for Efficient Churn Prediction in Telecom Industry. International Journal of Database Theory and Application, 9(8), pp.211-232.

[5] Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley, (2012) Customer churn prediction in telecommunications, Elsevier, Expert Systems with Applications: An International Journal, 39 (1), pp.1414-1425. https://linkinghub.elsevier.com/retrieve/pii/S0957417411011353

[6] Michael C. Mozer, Richard Wolniewicz, David B. Grimes, Eric Johnson, and Howard Kaushansky, (2020) Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. IEEE Transactions on Neural Networks 11(3), 690.

[7] Emiliano G. Castro and Marcos S. G. Tsuzuki, (2015) Churn Prediction in Online Games Using Players' Login Records: a Frequency Analysis Approach, IEEE Transactions on Computational Intelligence and AI in Games, IEEE Transactions on Computational Intelligence and AI in Games, 7(3), pp.255-265.

[8] Ning Lu, Hua Lin, Jie Lu, (2014) Member, IEEE, and Guangquan Zhang. A Customer Churn Prediction Model in Telecom Industry Using Boosting. IEEE Transactions on Industrial Informatics, 10(2), pp. 1659-1665.

[9] Marcin Bienkowski, Nadi Sarrar, Stefan Schmid, and Steve Uhlig, (2018) Online Aggregation of the Forwarding Information Base: Accounting for Locality and Churn, IEEE/ACM Transactions on Networking, 26(1), pp.591- 604.

[10] Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley, (2012) Customer churn prediction in telecommunications, Expert Systems with Applications, 39, pp: 1414–1425.

[11] Sven F. Crone, Stefan Lessmann, Robert Stahlbock, (2006) The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing, European Journal of Operational Research, 173, pp.781–800.

[12] Inderpreet Singh a, Sukhpal Singh, (2017) Framework for Targeting High Value Customers and Potential Churn Customers in Telecom using Big Data Analytics, I.J. Education and Management Engineering, 7(1), pp.36-45.

[13] Yiqing Huang, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, Wenyuan Dai, Qiang Yang Jia Zeng, (2015) Telco Churn Prediction with Big Data, SIGMOD'15, pp. 607-618.

[14] Adnan Amin, Changez Khan, Saeed Shehzad, Imtiaz Ali, Sajid Anwar, (2015) Churn Prediction in Telecommunication Industry Using Rough Set Approach, New Trends in Computational Collective Intelligence, 572, pp. 83- 95.

[15] Abdelrahim Kasem Ahmad, Assef Jafar and Kadan Aljoumaa, (2019) Customer churn prediction in telecom using machine learning in big data platform, Journal of Big Data, 6:28, pp.1-24.

[16] Dr. T. Arumuga Maria Devi and S. Mariammal, (2016) SVM Based Performance of IRIS Detection, Segmentation, Normalization, Classification and Authentication Using Histogram Morphological Techniques, International Journal of Computer Engineering and Technology, 7(4), pp. 1–11

[17] V. Muralidharan, V.Sugumaran, Gaurav Pandey, (2011) SVM Based Fault Diagnosis of Monoblock Centrifugal Pump Using Stationary Wavelet Features, International Journal of Design and Manufacturing Technology, 1 (2), pp. 1–6

[18] Sandip S. Patil and Asha P. Chaudhar, (2012) Classification of Emotions from Text Using SVM Based Opinion Mining, International Journal of Computer Engineering and Technology, 3(1), pp. 330–338