CSE6006 – NoSQL Databases

J Component - Project Report

Customer Churn Prediction

By

20MCS1002	P.ATHIKA
20MCS1012	KAZI ALTAMESH
20MCS1001	R.SWATHI SRI

M.Tech CSE

Submitted to

Dr. A. Bhuvaneswari,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

ABSTRACT

A CRM lets you store information about customers to help you track the status of every customer relationship. This can help businesses keep track of their clients and ultimately increase sales. The application will be able to store and edit customer details, as well as keep notes about them. CRM is a tool that allows companies to manage everything to-do with their customers. machine-learning algorithms had been implemented for the purpose of optimally predicting the possible churning customers and making the right decisions at the right moments. Researchers had conducted several studies on various types of algorithms and results were found very promising.

TABLE OF CONTENTS

I. Introduction

1.1 Objective and goal of the project

A. Improved Customer Satisfaction

Happy customers are loyal customers and they also offer good word of mouth advertising, which can be invaluable. Making customer satisfaction the primary goal of your CRM campaign is the surest way to improve your bottom line. This can be achieved by fostering increased customer engagement via social networking sites and various mobile platforms.

B. Improve the Efficiency Of Your Business

CRM can help you eliminate redundancies in your marketing campaigns by allowing you to intuit which stage of the purchasing process each returning customer is in. You can send out marketing materials that are targeted to specific interests and purchasing abilities, rather than transmitting general messages that are far less likely to generate an optimal amount of attention. A good CRM platform will collect and organize a wealth of data pertaining to individual and group consumer profiles.

C. Expand You Customer Base

It is important to note that CRM is not just for managing the customers you already have. A CRM program that is linked to a high-quality knowledge management platform will allow you to stay in contact with prospects that have yet to convert. It will also allow you to identify commonalities and relationships among the clients that you already have, so that you can hone and improve your future efforts in outreach. A larger customer base will allow for increased continuity in profits, even in a seasonal industry.

D. Enhance Your Sales and Support Teams

One of the most important CRM objectives to consider is enhancing your team. More importantly, CRM will give you a guideline for any professional development plans that you wish to implement and make mandatory among all employees. The data that your CRM system collects can be used as part of employee reviews and employee rewards or incentive programs. Tracking customer feedback in relation to the workers who have supplied customer service will allow you to make highly informed hiring, firing and promotion decisions. It will also give you an in-depth guide for sharing with any recruitment agencies that you are working with or for bolstering your own in-house recruitment teams.

1.2 Motivation

1) To assess factors which can improve customer relation in
2) To examine the advantages of effective management of customer relationship in
3) To investigate benefits of using customer retention approach in
4) To examine how companies can achieve good customer relationship.

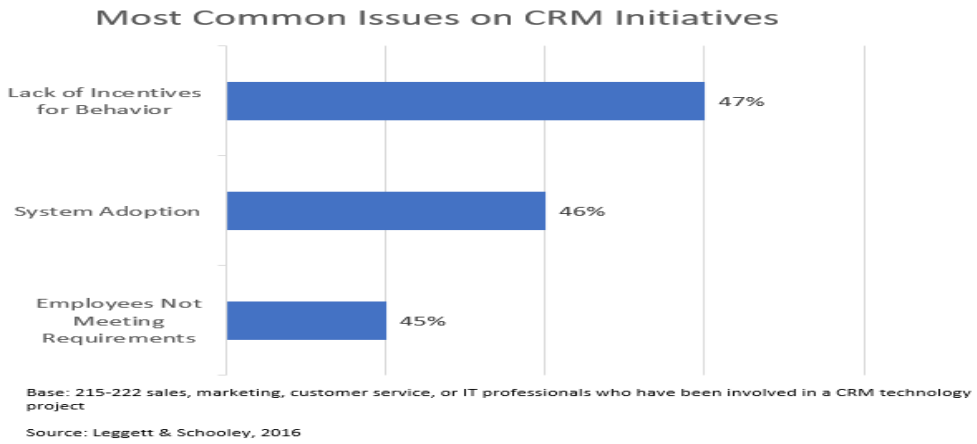
1.3 Challenges

In fact, the top 3 reported issues with CRM initiatives might all be people problems:

1) Lack of rewards for customer-oriented behavior (47%),
2) System adoption issues (46%),
3) Employees not meeting business and technology requirements (45%).

1.4 Contribution:

understanding the end goal of our project.Since our project has a defined ending.each contributor has desired end results. Swathi Sri.R(20MCS1001)has worked with UI and connectivity with databases.Athika P(20MCS1002) has worked with Data exploratory from given predicted content.Altamash kazi(20MCS1012) has worked on Machine learning after prediction of features.



II. Aim of project:

The aim of this project is to study some of the most important churn prediction techniques developed over the recent years. The primary objective is on the churn in telecom industries to accurately estimate the customer survival and customer hazard functions to gain the complete knowledge of churn over the customer tenure. Another objective is the identification of the customers who are at the blink of churn and approximating the time they will churn. This focuses on analyzing the churn prediction techniques to identify the churn behavior and validate the reasons for customer churn. This summarizes the churn prediction techniques in order to have a deeper understanding of the customer churn and it shows that most accurate churn prediction is given by the hybrid models rather than single algorithms so that telecom industries become aware of the needs of high risk customers and enhance their services to overturn the churn decision.

III. Literature Survey

A. Prediction of collaborative relationships by using network representation learning:

As the importance of supplier-customer relationships has become increasingly apparent that guides modern research and practice, the main impact of such researches is poured into the field of business management and operation research. However, prior

studies have indicated that firms tended to manage their relationships in a more structural and relational approach, no existing literature have applied NRL into predicting business relationships [1].

B. Using an opinion mining approach to exploit Web content in order to improve customer relationship management:

A traditional market survey typically seeks out customers' opinions using voluntary questionnaires or focus group interviews. The surveys' perceived and actual reliability could be limited by the number of customers who choose to respond, bias inherent in the wording of the questions or the subjects' interpretation regarding the information being sought. Online opinion resources such as review sites, forums, discussion groups and blogs are available with increasing variety and popularity. Companies can now apply information retrieval, natural language processing and machine learning techniques to automatically and more objectively identify and understand the opinions of their customers [2].

C. A comparison of machine learning techniques for customer churn prediction

This work constitutes a comparison of five of the most widely used classification methods on the problem of customers' churning in the telecommunication sector. In particular, we compare the performance of multi-layer Artificial Neural Networks, Decision Trees, Support Vector Machines, Naïve Bayes classifiers, and Logistic Regression classifiers, compared to their boosting versions in an attempt to further improve their performance [3].

D. Customer Satisfaction and Customer Perceived Value and its Impact on Customer Loyalty:

The Mediational role of customer relationship management.In the present-day, many researchers speculate that service related component which have immensely prestigious effect on decision-making measures for the customer; these factors could be quality of the service, customer satisfaction and customer value .Nonetheless, by various researchers relationship between customer satisfaction and loyalty has been highlighted questionably that firms should not entirely rely on customer satisfaction as a key factor and exclusive antecedent for customer loyalty [4].

E. Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector:

Factors of potential churn customer's usage and behavior which can later be used in customer profiling to specify policies for retention. There exist other methods for rule generation such as Rough Set Theory (RST) [33]. Rough Set Decision based Tree (RDT) performs well, however, in this study we performed customer profiling based on

their behavior through *k-means* clustering algorithm for creating retention policies by decision makers [5].

F. A survey on customer relationship management:

For every business success rely upon customers and there arise need to know about pulse of every customers. For that enterprise should know the knowledge about customers' patterns and behavior. With CRM helps to get in-depth knowledge of customer's pattern and their behavior which in turn applied in business to make more successful. Analytical techniques are applied in CRM customer data to discover knowledge and patterns. Further how customer's data privacy is held should be addressed. Further use of big data technologies in CRM can be addressed and use of different data mining techniques can be applied in customer data [6].

G. The Moderation Influence of Power Distance on the Relationship Between Technological Factors and the Successful Implementation of Citizen Relationship Management in the Public Sector:

This study found that there is a direct and significant impact of IT infrastructure and system quality on SICzRM. The finding of this study also showed that power distance has a moderation effect on the relationship between IT infrastructure and system quality with SICzRM. But it has no direct effect over SICzRM. This finding confirmed that power distance as a national culture dimension can be added to the TOE theory as a moderator [7].

H. A Survey on Customer Churn Prediction using Machine Learning Techniques:

This paper provides a detailed study on the methods used for the process of customer churn prediction. Each of the above churn prediction models has low accuracy and prediction. Hence a good prediction model is required in order to avoid the customer churn problem. Good prediction models have to be constantly developed and a combination of the proposed methods has to be used [8].

I. A Multi-Layer Perceptron Approach for Customer Churn Prediction

Data mining technique refers to extracting latent, unknown, meaningful and useful data information from large data sets to investigate the information that can be produced from the extracted data. In Europe, the number of customers who change operators has increased from year to year and the churn rate now stands at 25% up to 2012 on average. The loss of valuable customers will have impact on higher costs to

attract new customers, which is five to six times more expensive than customer retention expenses [9].

J. A Development of Web-based Customer Relationship Management (CRM) system for Beauty Clinic

A software development methodology (or what is called a process model or software engineering paradigm) is a process strategy that combines the processes, methods, and tools that build software in relation to a broad set of tasks involving requirements analysis, program construction, design, testing, and maintenance. In this development, the authors use the FAST model. By using the FAST method, the author becomes easier in terms of collecting the information needed for the system requirements created in the Almeera SkinCare Beauty Clinic Service Information System [10].

IV. Problem Statement

The customer relationship is a key to the success of any company. Hence, a company that does to have effective client relationship management suffer from several severe competition disadvantages in the marketplace. One of the major disadvantages is that the company will lose the competitive edge that good customer relationship management provides. Another problem is that the company will not be able to have the product that addresses the needs of the buyers due to lack of appropriate customer relationship management. Moreover, the organization that fails to properly implement the effective customer relationship management strategy in the required period will not be effective in improvement of the services of the buyers as well as the services innovation which entirely rely on the investments. Also, the telecommunication companies, e-commerce, and big data might not be efficiently executed in time in order to recover from the situations of lack of proper management of customer relations with the company. Therefore, the company target should be focused on finding out issues concerning the relationship that exists between the corporation and their buyers. Also, the organization should evaluate the ways of improving an effective client's relationship with the company.

V. Existing System

Using an opinion mining approach to exploit Web content in order to improve customer relationship management:

Typically, a traditional market survey uses voluntary questionnaires or focus group interviews to look for client feedback. The quantity of consumers who decide to answer the surveys, the inherence of the formulation of questions or the interpreting by the subjects of the information being sought could restrict the perceptibility and real reliability of those surveys. There are more diverse and popular online opinion resources like review sites, forums, discussion groups and blogs. Companies are now entitled to

automatically identify and understand the thoughts of their customers through information retrieval, natural language processing and machines. [2].

Disadvantage:

In fact, the top 3 reported issues with CRM initiatives might all be people problems:

1) Lack of rewards for customer-oriented behaviour (47%),
2) System adoption issues (46%),
3) Employees not meeting business and technology requirements (45%).

VI. Proposed system:

Customer churn is a crucial activity in rapidly growing and ma-true competitive telecommunication sector and is one of the greatest importance for a project manager. Due to the high cost of acquiring new customers, customer churn prediction has emerged as an indispensable part of telecom sectors 'strategic decision making and planning process. It is important to forecast customer churn behavior in order to retain those customers that will churn or possible may churn. This study is another attempt which makes use of rough set theory, a rule-based decision making technique, to extract rules for churn pre-diction. Experiments were performed to explore the performance of four differ-ent algorithms (Exhaustive, Genetic, Covering, and LEM2). It is observed thatrough set classification based on genetic algorithm, rules generation yields mostsuitable performance out of the four rules generation algorithms. Moreover, byapplying the proposed technique on publicly available dataset, the results showthat the proposed technique can fully predict all those customers that will churnor possibly may churn and also provides useful information to strategic decisionmakers as well.

Customer churn is one of the mounting issues of today's rapidly growing and compet-itive telecom sector. The focus of the telecom sector has shifted from acquiring newcustomer to retaining existing customers because of the associated high cost [1]. Theretention of existing customers also leads to improved sales and reduced marketingcost as compared to new customers. These facts have ultimately resulted in customerchurn, prediction activity to be an indispensable part of telecom sector's strategicdecision making and planning process. Customer retention is one of the main objec-tives of CRM (customer relationship management). Its importance has led to the de-velopment of various tools that support some important tasks in predictive modellingand classification. In recent decades, the organizations are increasingly focusing on long term rela-tionships with their customers and observing a customer's behavior from time to time using various applied knowledge discovery (KDD) techniques to extract hidden relationships between the entities and attributes in a flood of data bank. These facts have attracted many companies to invest in CRM to maintain customers'information. Customer centric approach is very common, particularly in telecommu-nication sector to predict the customers' behavior based on historical data stored in CRM. Data maintained in such CRM systems can be converted into meaningful infor-mation to consider the mounting issue of customer churn to

identify the customer's churn activities before the customers are lost which increase the customer strength

Advantage:

We can share our experiences, we hope to spur additional interest and innovation in this exciting space.

Disadvantage:

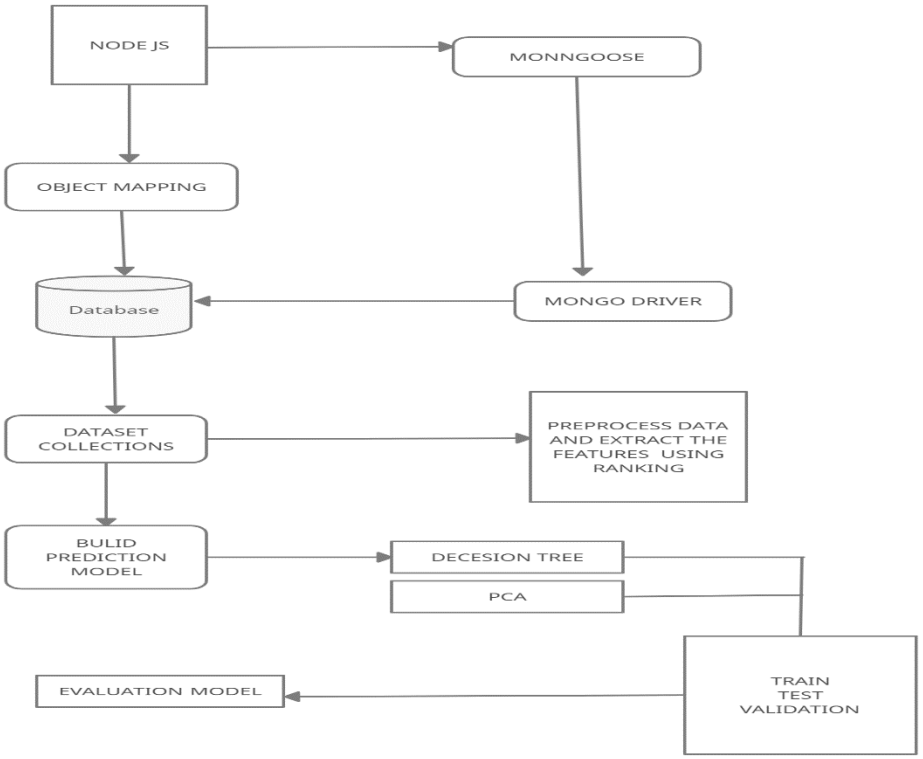
It requiring the knowledge of the social network and then modelling the diffusion.

VII. Requirements Specification

VII.1 Software Requirements

Google colab and node js.

VIII. System design

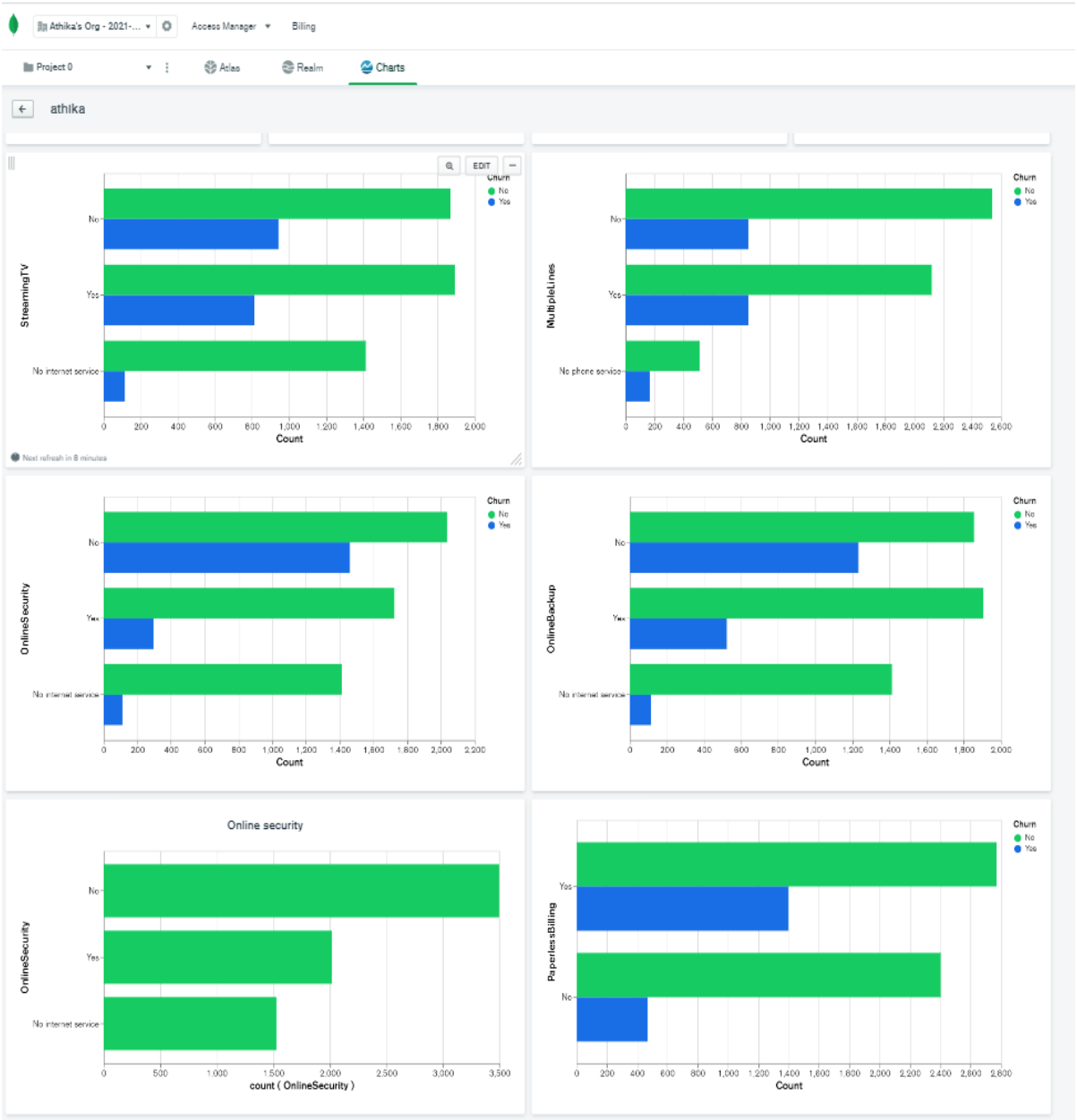


IX. Methodology

Mongodb schema representation for feature selection:

A classic market survey uses voluntary surveys or focus group interviews to search for customer input. The number of customers who opt to reply to surveys, the fact that questions are formulated or the subject matter interpreted might reduce the perceptibility and genuine dependability of the surveys. Online publications such as

review sites, forums, discussion groups and blogs are more diverse and popular. Companies now have the right to identify and comprehend their consumers' views automatically by collecting information, parsing natural languages and machines



Decision tree:

When customers discontinue doing business with a company, customer churn also called customer attrition. They are looking to discover these client groupings since the cost of gaining a new client is normally higher than that of retaining the previous customer. Using Telco Customer Churn information, we will construct a simple customer churn prediction model. We have used a decision tree to represent the customers churned, pandas for crushing data and matplots for viewing. We're going to do everything in Python above. With another dataset, the Code can be used with some

small modifications in order to train the basic model. We also share some information and ideas for additional features and upgrades. Variants of decision data have been implemented in SPSS data mining tool. A. Decision Tree A decision tree is a classification scheme which generates a flow chart like structure where an internal node represents a test on an attribute, each branch represents outcome of the test and leaf node represents classes. Decision tree partitions the input space into cells where each cell belongs to one class .

The decision tree is developed into two phases: building and pruning. In the building phase, data set partitioning is done till the records in a single partition contain identical values. On the other hand, in the second phase branches containing noisy data are removed. Decision tree is a nonparametric approach for building classification models. In other words, it does not require any prior assumptions regarding the type of probability distributions satisfied by the class and other attributes. At the same time, decision trees are relatively easy to interpret and are quit robust to the presence of noise especially the methods used for avoiding over-fitting. As decision tree is a powerful tool for classification and prediction by finding out the underlying and important patterns or relationships among data, it is one of the most frequently used data mining methods [13-17].

Classification means to divide subjects into some classes based on several predictor variables, for example, classing customers as churn or nonchurn according to their previous call details. Prediction means to forecast a future event based on rules discovered from a pool of data. Decision tree is made up in the form of a tree built by making child-nodes until each branch reaches the terminal node. Decision tree has three types of nodes: root node, internal nodes, and leaf or terminal nodes. In a decision tree, each leaf node is assigned a class label. The non-terminal nodes, which include the root and other internal nodes, contain attribute test conditions to separate records that have different characteristics.

X. Result analysis

```
[ ] xr_train,xr_test,yr_train,yr_test=train_test_split(X_resampled, y_resampled,test_size=0.2)

model_dt_smote=DecisionTreeClassifier(criterion = "gini",random_state = 100,max_depth=6, min_samples_leaf=8)

model_dt_smote.fit(xr_train,yr_train)
yr_predict = model_dt_smote.predict(xr_test)
model_score_r = model_dt_smote.score(xr_test, yr_test)
print(model_score_r)
print(metrics.classification_report(yr_test, yr_predict))
```

```
0.926131850675139
              precision    recall  f1-score   support

           0       0.91      0.92      0.92       542
           1       0.94      0.93      0.93       717

    accuracy                           0.93      1259
   macro avg       0.92      0.93      0.92      1259
weighted avg       0.93      0.93      0.93      1259
```

XI. Conclusion

The importance of the types of research in the telecom market is to help companies to make profit. It has become known that predicting churn is one of the most important source of income to telecom companies. Hence this research aimed to build a system that predicts the churn of customers

XII. References

1. Y. Zuo and Y. Kajikawa, "Prediction of collaborative relationships by using network representation learning," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 2017, pp. 69-74, doi: 10.1109/SMC.2017.8122580.
2. *A. Kongthon, N. Angkawattanawit, C. Sangkeettrakarn, P. Palingoon and C. Haruechaiyasak, "Using an opinion mining approach to exploit Web content in order to improve customer relationship management," PICMET 2010 TECHNOLOGY MANAGEMENT FOR GLOBAL ECONOMIC GROWTH, Phuket, Thailand, 2010, pp. 1-6.*
3. M. Singh, S. Singh, N. Seen, S. Kaushal and H. Kumar, "Comparison of learning techniques for prediction of customer churn in telecommunication," 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), Sydney, NSW, Australia, 2018, pp. 1-5, doi: 10.1109/ATNAC.2018.8615326.
4. M. C. O. Mendoza, R. R. C. Santos and J. E. H. Magdaraog, "Assessment of E-Service Quality Dimensions and Its Influence on Customer Satisfaction: A Study on the Online Banking Services in the Philippines," 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), Bangkok, Thailand, 2020, pp. 1076-1081, doi: 10.1109/ICIEA49774.2020.9101940.
5. I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

6. D. Prabha and R. S. Subramanian, "A survey on customer relationship management," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICACCS.2017.8014601.
7. S. D. Alabri, S. Kamaruddin, A. R. Gilal, J. Jaafar and I. A. Aziz, "The Moderation Influence of Power Distance on the Relationship Between Technological Factors and the Successful Implementation of Citizen Relationship Management in the Public Sector," in IEEE Access, vol. 8, pp. 132446-132465, 2020, doi: 10.1109/ACCESS.2020.3010126.
8. Y. Gan and D. Li, "The impact of the customer satisfaction, switching costs and trust on customer relationship commitment," 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, Xi'an, China, 2013, pp. 189-192, doi: 10.1109/ICIII.2013.6703545.
9. Li Xin and W. Rong, "Survey Research on Relationship among Service Failures, Service Recovery and Customer Satisfaction," 2007 International Conference on Management Science and Engineering, Harbin, China, 2007, pp. 1121-1126, doi: 10.1109/ICMSE.2007.4421996.
10. Sarwindah, Marini, Hengki and S. Febriyanti, "An Development of Web-based Customer Relationship Management (CRM) system for Beauty Clinic," 2020 8th International Conference on Cyber and IT Service Management (CITSM), Pangkal, Indonesia, 2020, pp. 1-5, doi: 10.1109/CITSM50537.2020.9268918.