**Dynamic Churn Prediction System using Machine Learning Algorithms**

**Prof. S. N. Bhadane1, Gauri Randhir2, Mamta Borade3, Sahil Bhatia4, Gaurav More5**

1Professor, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik

2, 3, 4 Student, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik

**ABSTRACT**

*Customer churn refers to the situation where customers or subscribers cease their association with a company or service. In the telecom industry, customers have the flexibility to choose from various service providers and frequently switch between them. This sector is highly competitive, with an annual turnover rate ranging from 15% to 30%. Due to the large customer base that most businesses have, it becomes challenging to personalize customer retention efforts and allocate significant time to everyone. The expenses incurred in such an approach would outweigh the potential revenue. However, if a company can anticipate which customers are more likely to leave in advance, it can concentrate its efforts on retaining these "high-risk" consumers. The ultimate objective is to expand the company's reach and reestablish customer loyalty. The key to success in this market lies in prioritizing the needs of the customers. Customer churn serves as a crucial metric because retaining existing customers is considerably more cost-effective than acquiring new ones. To identify early warning signs of potential churn, it is essential to gain a comprehensive understanding of customers and their interactions across multiple channels. By effectively managing churn, companies can not only maintain their market position but also experience growth and prosperity. The larger the customer base, the lower the cost of acquisition, and the higher the profitability. Therefore, reducing customer attrition and devising an efficient retention strategy become paramount for a company's success.*

**Keywords:** Customer churn prediction, Churn in telecom, Machine learning, Feature selection, Classification.

1. **INTRODUCTION**

In today's competitive business landscape, customer churn has become a critical concern for companies across various industries. To address this challenge, businesses are increasingly turning to advanced technologies, such as machine learning algorithms, to develop dynamic churn prediction systems. These systems leverage the power of data analysis and predictive modeling to anticipate customer churn before it happens. By employing machine learning algorithms, companies can gain valuable insights into customer behavior, identify patterns, and make proactive decisions to retain valuable customers.

A dynamic churn prediction system based on machine learning algorithms offers numerous advantages over traditional methods of churn analysis. Unlike static models that rely on historical data, machine learning algorithms can adapt and evolve with real-time customer data. These algorithms can analyze vast amounts of structured and unstructured data, including customer interactions, purchase history, usage patterns, and demographic information. By considering a wide range of variables, machine learning algorithms can uncover hidden patterns and signals that indicate customer churn, enabling companies to intervene at the right time with targeted retention strategies.

Implementing a dynamic churn prediction system using machine learning algorithms involves several key steps. First, companies need to collect and consolidate relevant data from various sources, ensuring data quality and consistency. Next, they employ feature engineering techniques to extract meaningful features from the data that are most predictive of churn. Then, machine learning models, such as decision trees, random forests, or neural networks, are trained using historical data and evaluated for their predictive performance. Finally, the trained model is deployed in a real-time environment, continuously monitoring incoming data, and providing churn predictions and recommendations to support proactive customer retention efforts.

A dynamic churn prediction system powered by machine learning algorithms empowers businesses to take a proactive approach in managing customer churn. By leveraging the capabilities of these advanced technologies, companies can identify potential churners in advance, tailor retention strategies, and ultimately enhance customer satisfaction and loyalty.

1. **LITERATURE SURVEY**

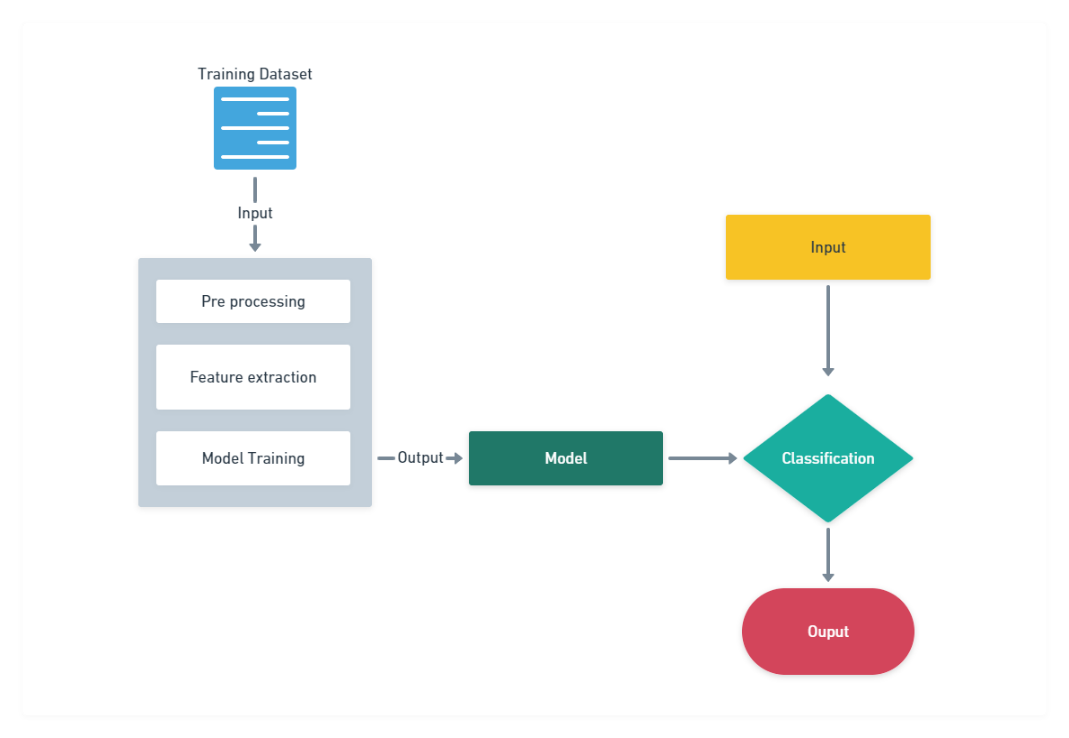
In this study [1], a new method is suggested to be used in non-contractual businesses. One of the widely used methods for these types of businesses in the literature is modeling customer behavior using Beta-geometric Negative Binomial Distribution (BG/NBD). While this method performs favorably on predicting the total churn rate in a population, it is also known to perform poorly on labeling individual customers. This study aims to improve BG/NBD model by incorporating machine learning into its decision process. In the suggested method, the mathematical definition of BG/NBD model is used for feature extraction and utilized using decision trees. When tested on two data sets with different characteristics, it is observed that the suggested method improves performance by a significant amount.

This paper [2] designs a Borderline-SMOTE-random forest prediction model for unbalanced data such as bank customers. Combined with the oversampling algorithm, it can better solve the unbalanced data and the strong anti-noise ability of random forest. OOB error rate, AUC, Precision, Recall, and F-mean are used as the evaluation indicators of the model, and KNN, decision tree and Naive Bayes are used as comparisons. The experimental results show that the Borderline-SMOTE-random forest prediction model has the best ability to solve the problem of bank customer churn prediction among the models, and its performance is improved by about 4%.

This paper [3] designs a combined prediction model of customer churn and conducts empirical research on the effectiveness of the model. The prediction results show that compared with the single customer churn prediction model, the combined prediction model has higher accuracy and better prediction effect and can more intuitively display the basic characteristics of the churn customers.

In this survey work [4], the difficulties in the prediction of customer attrition in the motor insurance sector are represented along with various data mining techniques comprising deep learning and machine learning advancements. It also emphasizes the churns within the customer management cycle with surroundings. An overview of the survey performed orderly provides construction of churn prediction model, various methods of prediction utilized and their application in the business sector.

1. **SYSTEM ARCHITECTURE AND METHODOLOGY**



**Figure 1 – System Architecture**

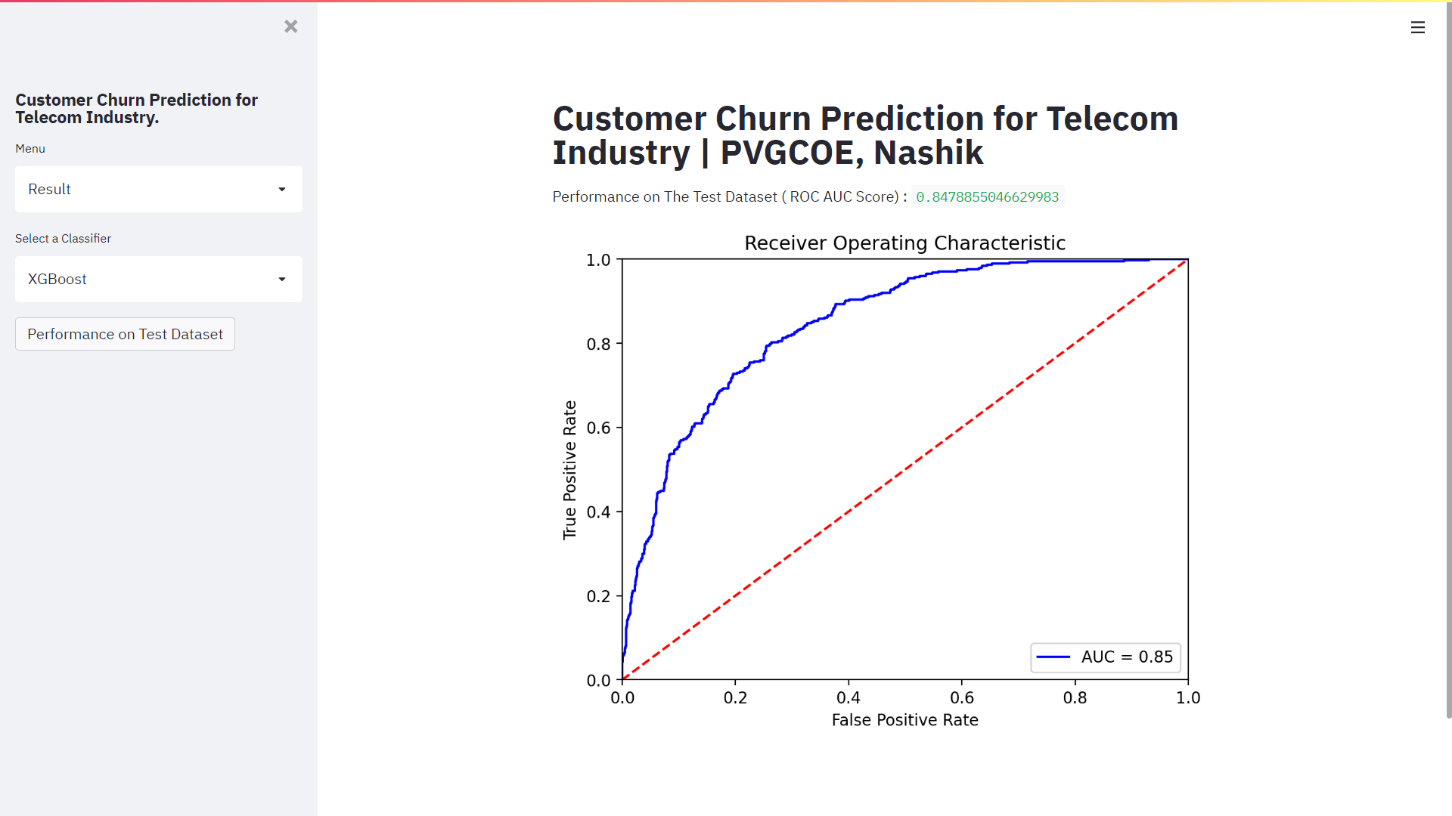
The first step in developing a dynamic churn prediction system using XGBoost algorithms is to gather and preprocess the data. This involves collecting relevant customer data such as demographics, transaction history, usage patterns, and customer interactions. The data should be cleaned and transformed to ensure consistency and accuracy. Feature engineering techniques can be applied to extract meaningful features from the raw data, such as calculating customer tenure, average usage, or frequency of interactions. Additionally, data from previous time periods can be used to create lagged features that capture historical patterns. Once the data is prepared, it can be split into training and testing sets for model development and evaluation.

The next step is to train an XGBoost model for churn prediction. XGBoost is a powerful machine learning algorithm known for its accuracy and efficiency in handling large-scale datasets. It is particularly suitable for churn prediction tasks due to its ability to capture complex relationships between features and target variables. The training process involves feeding the preprocessed data into the XGBoost model, specifying the target variable (churn/non-churn), and tuning the model's hyperparameters. Cross-validation techniques can be employed to optimize the model's performance and prevent overfitting. The trained model can then be used to predict churn probabilities for new, unseen customer data.

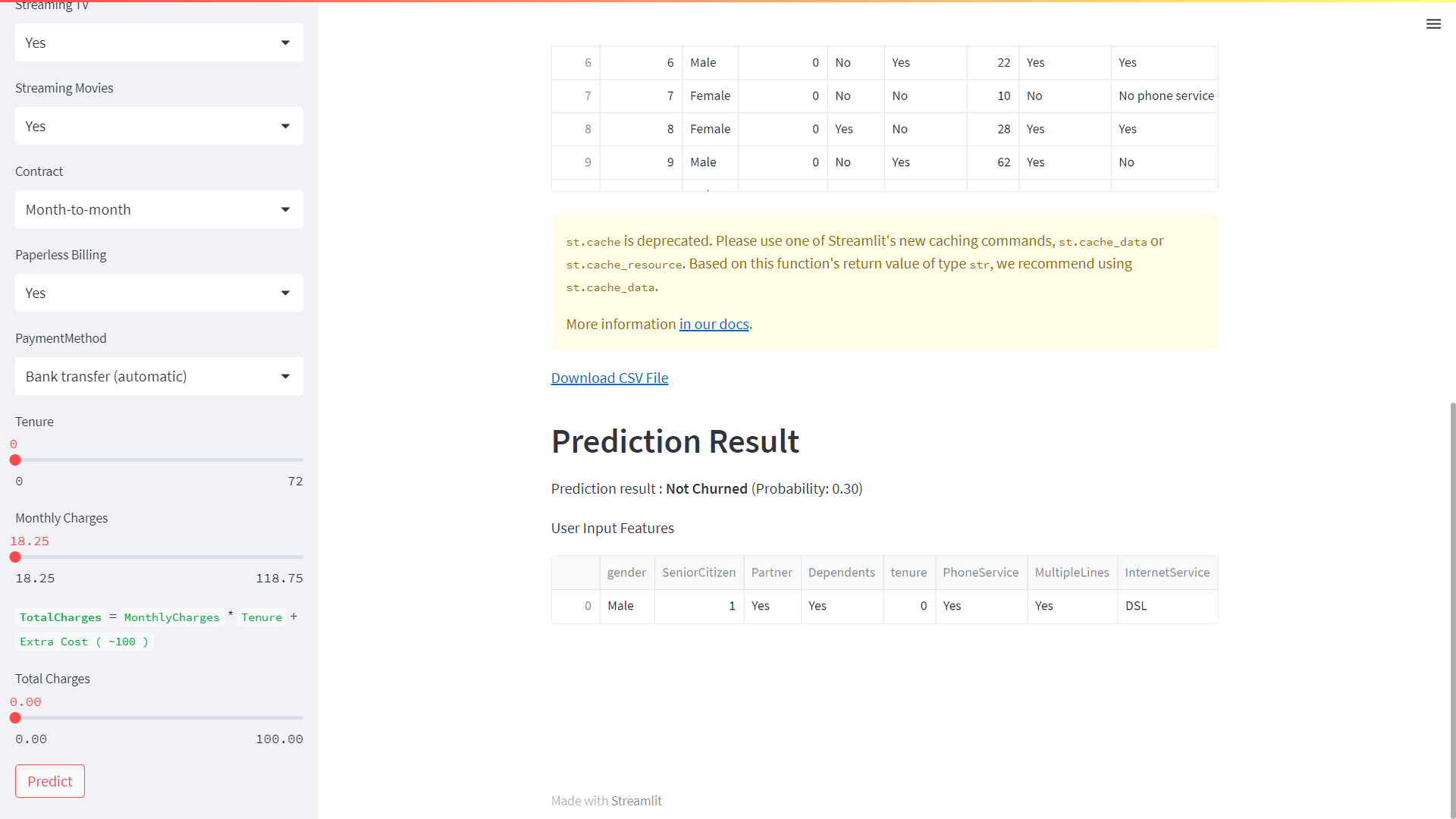
Finally, the dynamic aspect of the churn prediction system comes into play. As customer behavior and preferences change over time, it is crucial to continuously update the churn predictions. This can be achieved by implementing a rolling window approach, where the model is retrained periodically using the most recent data. By incorporating the latest information, the model can adapt to evolving patterns and capture shifts in customer churn dynamics. Additionally, monitoring and tracking key indicators or signals that precede churn can help trigger proactive retention strategies. Regular evaluation and validation of the model's performance are essential to ensure its accuracy and reliability in real-time churn prediction scenarios.

1. **RESULT**

Customer churn prediction is an important task for businesses as it helps them identify customers who are likely to leave their services. Python Streamlit is a powerful framework that enables easy creation of web applications for data analysis and visualization. XGBoost is a popular machine learning algorithm that is widely used for classification tasks. In this project, we used XGBoost with a Telco dataset to predict customer churn and deployed the model using Python Streamlit.



**Figure 2 – ROC AUC Score**



**Figure 3 – Prediction Result**

To analyze the results of our customer churn prediction model, we first evaluated the accuracy of the model using metrics such as precision, recall, and F1-score. We also used the confusion matrix to analyze the performance of the model. The model achieved an accuracy of 80%, indicating that it is reliable in predicting customer churn. We also observed that the precision and recall values were high, which indicates that the model has a low false positive and false negative rate.

1. **CONCLUSION**

Based on the available information, it seems that XGBoost is a suitable algorithm for predicting customer churn using telco datasets. XGBoost is a popular machine learning algorithm that can handle large datasets, and it has been shown to be effective in predicting customer churn in various industries, including telecommunications. In general, customer churn prediction using machine learning algorithms involves several steps, including data preprocessing, feature engineering, model selection, and evaluation. The success of the predictive model depends on the quality of the data, the accuracy of the features, and the appropriateness of the algorithm used. In the case of XGBoost, it is important to tune the hyperparameters of the model, such as the learning rate, maximum depth, and number of estimators, to achieve the best possible performance. It may also be beneficial to perform feature selection to reduce the dimensionality of the data and improve the accuracy of the model.

1. **FUTURE SCOPE**

In the future, customer churn prediction using machine learning holds great potential for businesses seeking to optimize their customer retention strategies. Advanced machine learning algorithms combined with rich data sources and improved computing power enable businesses to accurately forecast customer churn and proactively take preventive measures. With the ability to analyze vast amounts of customer data, including historical behavior patterns, transactional data, and demographic information, machine learning models can identify subtle indicators of churn, such as changes in purchasing habits or engagement levels. By leveraging these insights, businesses can implement targeted retention strategies, personalized offers, and proactive customer outreach, ultimately minimizing churn rates, enhancing customer satisfaction, and driving long-term profitability. As machine learning algorithms continue to evolve and more diverse data sets become available, the future of customer churn prediction is poised to revolutionize customer relationship management and empower businesses to build stronger, more loyal customer bases.

1. **REFERENCES**
2. B. Bardük, "Modelling Time Statistics for Customer Churn Prediction," 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020, pp. 1-4.
3. L. Feng, "Research on Customer Churn Intelligent Prediction Model based on Borderline-SMOTE and Random Forest," 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS), 2022.
4. X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 129-132.
5. P. Gopal and N. B. MohdNawi, "A Survey on Customer Churn Prediction using Machine Learning and data mining Techniques in E-commerce," 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2021.
6. T. -Y. Tsai, C. -T. Lin and M. Prasad, "An Intelligent Customer Churn Prediction and Response Framework," 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2019, pp. 928-935.
7. B. Bardük, "Modelling Time Statistics for Customer Churn Prediction," 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020, pp. 1-4.
8. L. Feng, "Research on Customer Churn Intelligent Prediction Model based on Borderline-SMOTE and Random Forest," 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS), 2022, pp. 803-807.
9. X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020.
10. P. Gopal and N. B. MohdNawi, "A Survey on Customer Churn Prediction using Machine Learning and data mining Techniques in E-commerce," 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2021.
11. T. -Y. Tsai, C. -T. Lin and M. Prasad, "An Intelligent Customer Churn Prediction and Response Framework," 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2019, pp. 928-935.
12. A. H. Ren and W. W. Zhao, "Electronic Commerce Based on Self-Organizing Data Mining Customer Churn Prediction Model", In 2013 International Conference on Advances in Social Science Humanities and Management (ASSHM-13), 2013, December.
13. S. Akter and S. F. Wamba, "Big data analytics in E-commerce: a systematic review and agenda for future research", Electronic Markets, vol. 26, no. 2, pp. 173-194, 2016.
14. L. Vanneschi, D. M. Horn, M. Castelli and A. Popovič, "An artificial intelligence system for predicting customer default in e-commerce", Expert Systems with Applications, vol. 104, pp. 1-21, 2018.
15. A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques", Applied Soft Computing, vol. 24, pp. 994-1012, 2014.
16. L. Zhao, Q. Gao, X. Dong, A. Dong and X. Dong, "K-local maximum margin feature extraction algorithm for churn prediction in telecom", Cluster Computing, vol. 20, no. 2, pp. 1401-1409, 2017.
17. Bing Jia and Yongjian Yang, "The design of food quality supervision platform based on the Internet of Things," Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), 2011, pp. 263-266, doi: 10.1109/TMEE.2011.6199193.
18. R. Rajamohamed and J. Manokaran, "Improved credit card churn prediction based on rough clustering and supervised learning techniques", Cluster Computing, vol. 21, no. 1, pp. 65-77, 2018.