IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.13

Explainable AI-Based Customer Churn Prediction, Retention Strategy for the-Commerce Industry

Prof.G.P.Raja
Associate Professor, Department of AI & Data Science
Sengunthar Engineering College (Autonomous), Tiruchengode, India
aidshod@scteng.co.in, gpraja1@gmail.com
https://orcid.org/0000-0002-5128-5312
Sridhar.S
UG Student, Department of AI & Data Science
Sengunthar Engineering College (Autonomous),
Tiruchengode, India
sridharsriram2005@gmail.com

Publication History
Manuscript Reference: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10092
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10092
Received: 30, January 2026, Revised: 13, February 2026, Accepted: 28 February 2026 Published Online: 25 March 2026
https://www.irjcs.com/volumes/Vol13/iss-03/13.CSMR26.MRCS10092.pdf
Article Citation: Prof.Raja,Sridhar(2026),Explainable AI-Based Customer Churn Prediction, Retention Strategy for the-Commerce Industry,IRJCS :International Research Journal of Computer Science, Volume 13,Issue 03 of 2026 pages 169-175 **Doi:>** https://doi.org/10.26562/irjcs.2026.v1303.13 **BibTeX Key Prof.Raja@2026Explainable**
Orcid: https://orcid.org/0009-0004-9398-7488
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1303.13 The journal's official abbreviation is IRJCS.

Abstract: Customer churn has become a significant challenge in the rapidly growing e-commerce industry, where customers frequently switch between online platforms due to competitive pricing, personalized services, and improved user experiences. Predicting potential churn customers in advance allows organizations to implement effective retention strategies and improve customer loyalty. This research proposes an Explainable Artificial Intelligence (XAI) based customer churn prediction system designed specifically for the e-commerce industry. The proposed framework utilizes machine learning algorithms such as Random Forest, Gradient Boosting, and Logistic Regression to analyze customer behavioral patterns including purchase history, browsing activity, transaction frequency, and customer engagement metrics. In addition to accurate churn prediction, explainable AI techniques such as SHAP and LIME are integrated to interpret model predictions and provide clear explanations for the factors influencing customer churn.
Keywords: Customer Churn Prediction, Explainable Artificial Intelligence, Machine Learning, E-Commerce Analytics, Customer Retention, SHAP, LIME.

INTRODUCTION

The rapid expansion of the global e-commerce market has significantly transformed the way customers interact with businesses. Online platforms offer consumers a wide range of products, competitive pricing, and convenient shopping experiences. However, the availability of multiple competing platforms has increased customer switching behavior, leading to higher churn rates. Customer churn refers to the phenomenon where customers stop purchasing products or discontinue using services from a particular platform. For e-commerce companies, customer retention is often more cost-effective than acquiring new customers. Studies indicate that retaining existing customers can be up to five times cheaper than acquiring new ones. Therefore, predicting potential churn customers in advance has become an essential task for business intelligence systems. Machine learning techniques have proven highly effective in identifying hidden patterns within large customer datasets. Algorithms such as Random Forest, Decision Trees, and Gradient Boosting can analyze complex behavioral data and predict customer churn with high accuracy. However, traditional machine learning models often operate as black boxes, making it difficult for businesses to understand the reasoning behind predictions. Explainable Artificial Intelligence (XAI) addresses this limitation by providing interpretable insights into machine learning models. By integrating explain ability techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), organizations can identify key factors influencing churn predictions and design effective customer retention strategies. This research proposes an Explainable-AI based customer churn prediction system designed specifically for the e-commerce industry. The system combines machine learning algorithms with explainable AI techniques to accurately identify potential churn customers while also providing interpretable insights into the prediction process. By integrating predictive analytics with retention strategy recommendations, the proposed system aims to support business decision-makers in improving customer engagement and reducing churn rates.

LITERATURE REVIEW

Customer churn prediction has attracted significant attention in recent years due to its impact on business profitability and customer relationship management.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 13, Issue 03, March 2026 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.13
Traditional statistical approaches such as logistic regression were initially used to predict churn by analyzing demographic and transactional data. However, these models often struggled to capture complex nonlinear relationships present in modern e-commerce datasets. Recent studies have demonstrated the effectiveness of machine learning algorithms such as Random Forest, Support Vector Machines (SVM), and Gradient Boosting in predicting churn. These algorithms can process large volumes of customer data and identify behavioral patterns associated with churn risk. Several research works have also explored the use of deep learning techniques for churn prediction. Neural network architectures can automatically extract meaningful features from customer activity logs and transaction histories, enabling improved predictive performance. Despite the advancements in predictive modeling, a major challenge remains the lack of interpretability in many machine learning systems. Business managers often require clear explanations of model predictions before implementing strategic decisions. Explainable AI techniques have emerged as a promising solution to this problem by providing transparent insights into the contribution of each feature. Techniques such as SHAP values and LIME explanations allow analysts to visualize feature importance and understand the reasons behind churn predictions. Integrating these methods with predictive models enables organizations to design targeted marketing campaigns, personalized offers, and customer engagement strategies. Several recent studies have explored the application of machine learning techniques for predicting customer churn in digital commerce environments. Researchers have demonstrated that predictive models built using historical customer data can significantly improve the identification of potential churn customers and enable proactive retention strategies. These models typically analyze behavioral attributes such as purchase frequency, browsing patterns, transaction history, and customer engagement metrics. Another line of research focuses on ensemble learning methods such as Random Forest and Gradient Boosting. These algorithms combine multiple decision trees to improve prediction accuracy and robustness. Studies have shown that ensemble methods outperform traditional statistical approaches like logistic regression in handling large and complex customer datasets commonly found in e-commerce platforms.

PROPOSED METHODOLOGY ARCHITECTURE

The proposed system architecture for the Explainable-AI based customer churn prediction system consists of multiple functional layers designed to process customer data, train predictive models, and generate interpretable insights. The first layer focuses on data acquisition, where customer data is collected from e-commerce platforms including transaction history, browsing behavior, product preferences, and engagement metrics. These datasets form the foundation for churn prediction analysis. The second layer performs data preprocessing and feature engineering. In this stage, missing values are handled, categorical variables are encoded, and irrelevant attributes are removed. Feature engineering techniques are applied to create meaningful variables such as purchase frequency, average order value, and customer lifetime value. The third layer consists of machine learning model training, where algorithms such as Random Forest and Gradient Boosting are trained on historical customer data. These models learn patterns associated with customer churn and classify users into churn or non- churn categories.

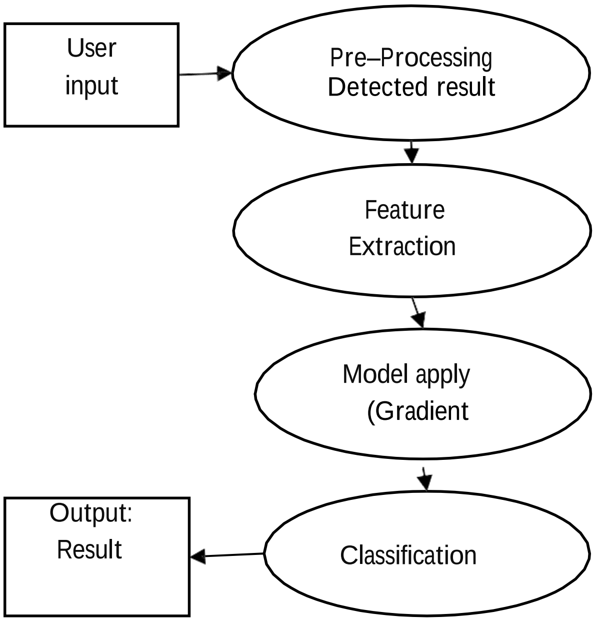

Fig.1. Architecture Diagram

A. Customer Behavior Monitoring

The system continuously monitors customer interactions within the e-commerce platform. Customer activity data includes purchase history, product browsing patterns, cart abandonment rate, transaction frequency, and customer engagement levels. These behavioral attributes help identify potential indicators of churn such as reduced purchase frequency, declining engagement, and inactive customer sessions. The monitoring system analyzes historical and real-time data to detect unusual customer behavior patterns that may indicate an increased risk of churn.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.13

B. Data Processing Framework

The data processing framework includes several stages such as data cleaning, normalization, and feature selection. Raw e-commerce datasets often contain missing values, duplicate records, or irrelevant attributes that need to be removed before analysis. Feature engineering techniques are applied to extract meaningful attributes from customer activity data. After cleaning, the next stage involves data normalization and transformation. Since customer attributes such as purchase amount, browsing duration, and transaction frequency may have different numerical ranges, normalization techniques are applied to scale the values into a consistent range. This process helps improve the performance of machine learning algorithms by preventing attributes with large values from dominating the model training process.

C. Churn Prediction Model

The churn prediction module utilizes machine learning algorithms to classify customers into churn and non-churn categories. Machine learning techniques are capable of identifying complex relationships between customer attributes and churn behavior. In this system, algorithms such as Random Forest, Logistic Regression, Decision Tree, and Gradient Boosting are used to analyze customer data. These algorithms are trained using historical datasets containing both churned and active customers. During the training process, the models learn patterns that indicate when customers are likely to stop using the platform. After training, the models analyze new incoming customer data and calculate churn probability scores for each customer. Customers with higher probability scores are identified as potential churn customers. These predictions allow organizations to focus their retention efforts on high-risk customers who are more likely to discontinue using the service

D. Explainable AI Module

To A key feature of the proposed system is the integration of explainable artificial intelligence techniques that provide transparency and interpretability for machine learning predictions. Traditional machine learning models often operate as black-box systems, making it difficult for business managers to understand the reasons behind the predictions. To address this issue, explainable AI techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are incorporated into the framework. These techniques analyze the contribution of each feature to the final prediction and provide detailed explanations for why a customer is predicted to churn. For example, the system may identify that low purchase frequency, reduced engagement, or high cart abandonment rate are key factors contributing to churn risk. These insights help business analysts understand customer behavior more clearly and design effective strategies to retain customers.

D. Retention Strategy Implementation

After identifying potential churn customers, the system recommends targeted retention strategies to reduce customer attrition. These strategies are designed tore-engage customers and encourage continued interaction with the e-commerce platform. Examples of retention techniques include personalized discount offers, loyalty reward programs, targeted email marketing campaigns, and product recommendation systems based on customer preferences

I. TECHNOLOGIES USED

A. Customer Data Collection Modules

The system collects customer data from multiple sources within the e-commerce platform, including transaction databases, browsing logs, and customer profile information. These data sources provide valuable insights into how customers interact with the platform. Important attributes such as purchase history, product categories viewed, browsing duration, transaction amounts, and customer engagement levels are collected and stored. This data forms the primary dataset used for machine learning-based churn prediction.

B. Data Storage and Processing

Customer data is stored in structured database systems and processed using modern data analytics tools. The collected data undergoes preprocessing steps such as cleaning, normalization, and feature transformation to prepare it for machine learning analysis. Efficient data processing techniques ensure that large volumes of customer activity data generated by e-commerce platforms can be handled effectively. This structured data environment enables accurate predictive modeling and supports further analytical operations.

C. Machine Learning Framework

Machine learning frameworks such as PyTorch, Tensor Flow, and Keras are used to build and train the APT detection models. These frameworks support the development of classification algorithms that analyze network flow features and identify malicious patterns. The trained models learn from historical traffic data and detect anomalies that indicate possible cyber threats, including data ex filtration, lateral movement, and command-and-control communication.

D. Churn Prediction Model

The churn prediction model is a core component of the proposed system, responsible for identifying customers who are likely to discontinue their interaction with thee-commerce platform. This module utilizes advanced machine learning algorithms to analyze customer behavioral data and classify users into churn and non- churn categories. Historical customer datasets containing both active and churned customers are used to train the predictive models.

E. Retention Strategy Implementation

Once the churn prediction model identifies customers who are likely to leave the platform, the retention strategy module is activated to reduce customer attrition. The primary objective of this module is to re-engage customers and encourage them to continue using the e-commerce platform. The system analyzes the reasons behind churn predictions using explainable AI techniques and recommends appropriate retention actions based on customer behavior patterns

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 13, Issue 03, March 2026 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.13

F. Backend Integration Framework

Explain ability plays a crucial role in ensuring that the predictions generated by the machine learning models are transparent and understandable for business decision-makers. Traditional predictive models often function as black-box systems, making it difficult for analysts to interpret the reasons behind churn predictions.

G. System Integration and Deployment

The proposed churn prediction system is designed to be integrated with existing e-commerce platforms and customer relationship management systems. The integration process involves connecting the data acquisition modules with transaction databases, customer activity logs, and online interaction records.

H. Performance evaluation

The effectiveness of the proposed churn prediction system is evaluated using several machine learning performance metrics that measure the accuracy and reliability of the predictive models. Metrics such as accuracy, precision, recall, and F1-score.

I. Advantages of the Proposed System

The proposed Explainable-AI based customer churn prediction system offers several advantages for e-commerce organizations. First, the system enables early identification of potential churn customers by analyzing customer behavior patterns and historical transaction data.

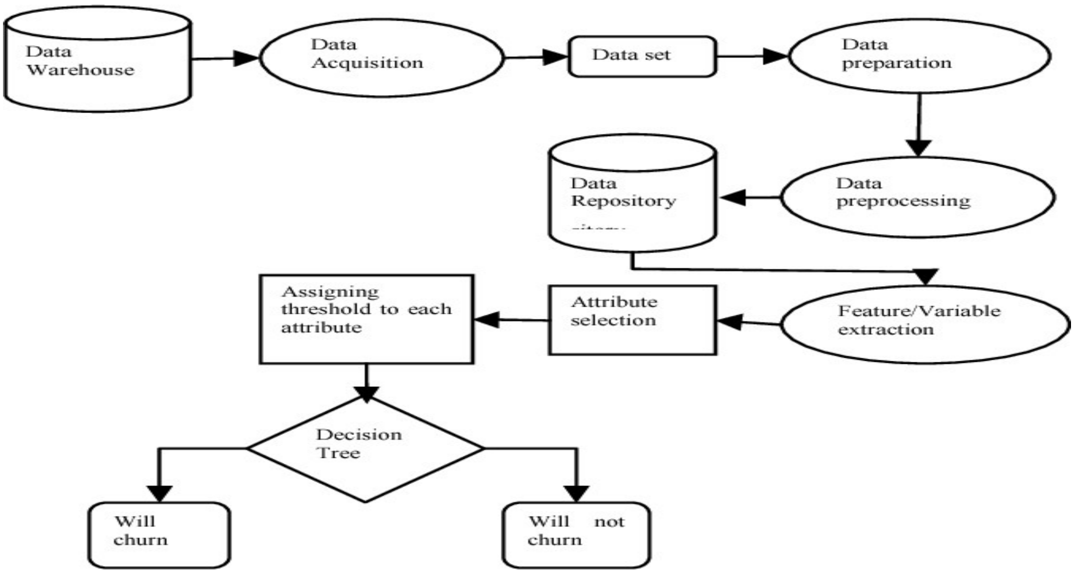

Fig.2: System Implementation

IMPLEMENTATIONS AND RESULTS

The system implementation focuses on developing a machine learning based framework capable of predicting customer churn and providing interpretable insights for thee-commerce industry. The implementation process begins with the collection of customer behavioral data from e-commerce platform. Fig.2.

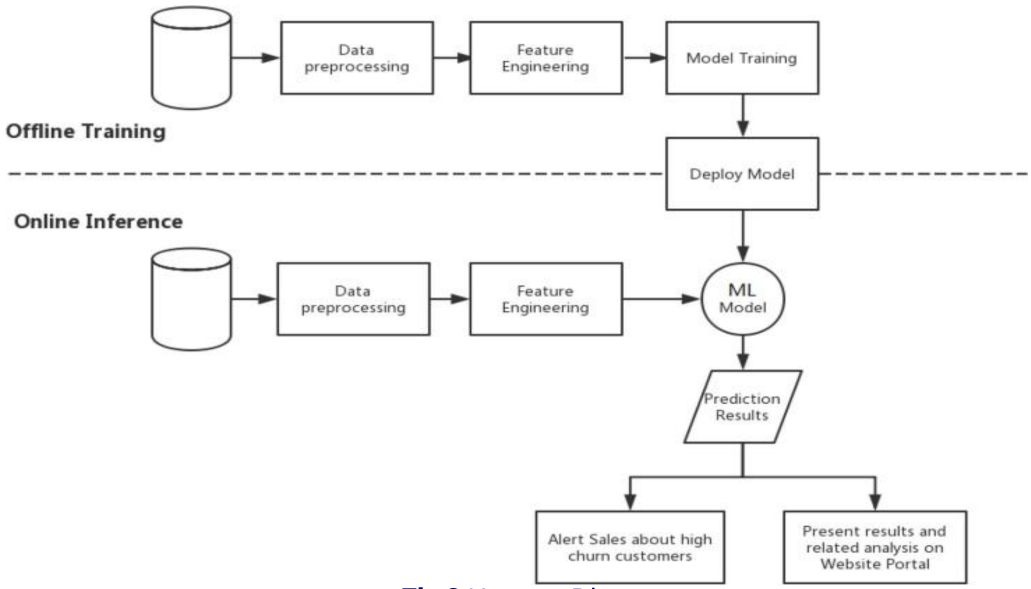

Fig.3 Use case Diagram

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.13

The use case diagram illustrates the interactions between the system and its primary users in the proposed customer churn prediction platform. The main actor in the system is the business analyst or administrator, who interacts with the system to manage customer data, perform churn prediction, and analyze results. The system allows the user to upload or access customer datasets that contain information such as purchase history, browsing behavior, and transaction records.

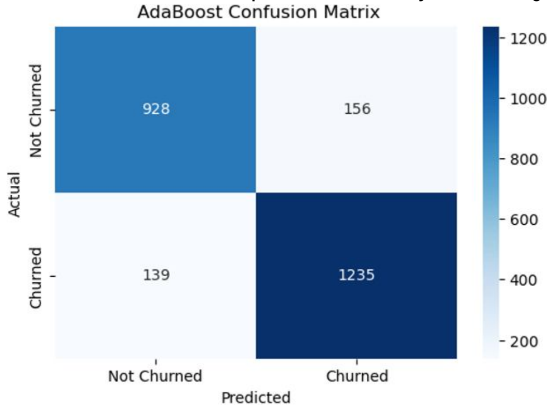

Fig.4 Customer Churn Confusion Matrix

The performance of the proposed customer churn prediction system is evaluated using a confusion matrix, as illustrated in Fig.4. The confusion matrix provides a detailed representation of the classification performance of the machine learning model by comparing the predicted outcomes with the actual customer churn labels.

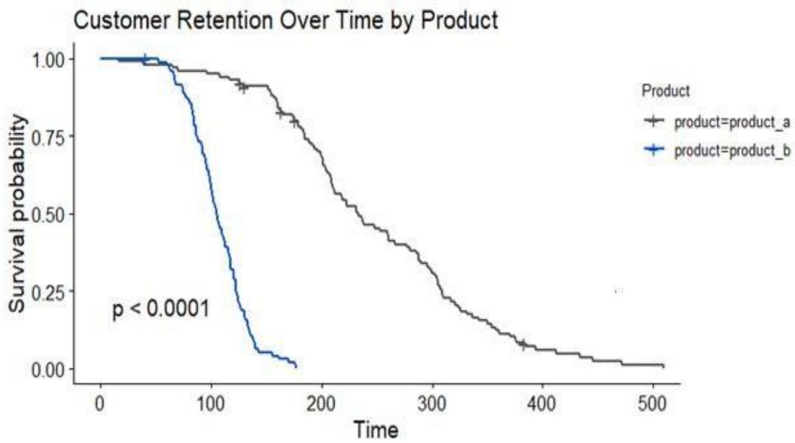

Fig.5 Customer Churn Distribution Graph

The distribution of predicted churn categories within the analyzed customer dataset is presented in Fig. 5.The graph illustrates the frequency of different customer categories such as active customers and churn customers based on the machine learning classification results. In most real-world e-commerce datasets, active customers typically dominate the dataset, while churn customers represent a smaller portion of the total population.

Key Considerations:

Data Quality and Availability: The accuracy and reliability of the churn prediction system depend heavily on the quality and availability of customer data collected from e-commerce platforms.

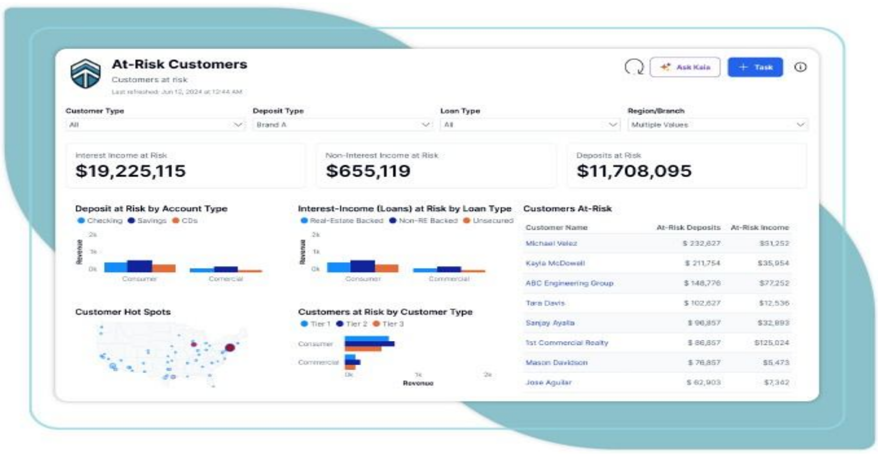

Fig 6: Prediction Page

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.13

Fig 7: Customer Retention Recommendation

Feature Selection: Selecting the most relevant features from the dataset is an important factor in improving churn prediction accuracy. Customer behavior attributes such as purchase frequency, transaction history, browsing duration, cart abandonment rate, and customer engagement level play a significant role in identifying churn patterns.

Model Accuracy and Performance: The machine learning algorithms used for churn prediction must be carefully optimized to achieve high accuracy while minimizing prediction errors. Performance evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure the effectiveness of the models

Real-Time Detection Capability: Customer behavior in e- commerce platforms changes continuously, and therefore the churn prediction system must be capable of analyzing customer data in near real-time.

VI. CONCLUSION

The proposed system presents an effective approach for predicting customer churn and supporting retention strategies in thee-commerce industry using machine learning and explainable artificial intelligence techniques. By analyzing customer behavioral data such as purchase history, browsing patterns, transaction frequency, and engagement levels, the system can accurately identify customers who are at risk of leaving the platform. The use of machine learning algorithms enables the system to automatically learn patterns from historical customer data and improve prediction accuracy. In addition, the integration of explainable AI techniques provides transparency in the prediction process by highlighting the key factors that influence churn behavior. This interpretability allows business analysts and decision-makers to better understand customer behavior and implement appropriate retention strategies such as personalized offers, loyalty programs, and targeted marketing campaigns

REFERENCES

1. H.Verbeke, D.Martens, C.Mues, and B.Baesens, “Building comprehensible customer churn prediction models with advanced rule induction techniques, ”Expert Systems with Applications,vol.38,no.3,pp.2354–2364,2011.
2. T.Coussement and D. Van den Poel, “Churn prediction in subscription services: An application of support vector machines, ”Expert Systems with Applications, vol.36,no. 3, pp. 4620–4627, 2009.
3. D.J.Huang, T.Kechadi, and B.Buckley, “Customer churn prediction in telecommunications,” Expert Systems with Applications, vol. 39, no. 1, pp. 1414–1425, 2012.
4. BA.Idris, A.Khan, and Y.S.Lee, “Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost based ensemble classification,” Applied Intelligence, vol. 39, no. 3, pp. 659–672, 2013.
5. H.S.Amin,A.Anwar,A.Adnan,M.Nawaz,K.Alawfi,A.Hussain, and K.Huang, “Customer churn prediction in the telecommunication sector using a rough set approach,” Neurocomputing, vol. 237, pp. 242–254, 2017.
6. R.G.Tsai and Y.Lu,“Customer churn prediction by hybrid neural networks,” Expert Systems with Applications, vol. 36, no. 10, pp. 12547–12553, 2009.
7. RX.Verbeke, K.Dejaeger, D.Martens, J.Hur, and B.Baesens, “New insights into churn prediction in the telecommunication sector,” European Journal of Operational Research, vol. 218, no. 1, pp. 211–229, 2012.
8. S.Moro,P.Cortez,andP.Rita,“Adata-driven approach to predict the success of bank telemarketing,” Decision Support Systems, vol. 62, pp. 22–31, 2014.
9. Q.R.Keramati and A.Ardabili, “Churn analysis for an Iranian mobile operator,” Telecommunications Policy, vol. 35, no. 4, pp. 344–356, 2011.
10. H.Bian, T.Bai, M.A.Salahuddin, N.Limam, A.A.Daya, and R.Boutaba, "Host in danger? Detecting network intrusions from authentication logs," in 2019 15th International Conference on Network and Service Management (CNSM), 2019, pp. 1-6.
11. O.McCusker, S.Brunza, and D.Dasgupta, "Deriving behavior primitives from aggregate network features using support vector machines," in 2013 5th International Conference on Cyber Conflict(CYCON2013),2013,pp.1-6.

12. Y.Xiuzhang, P.Guojun, L.Side, Z.Dongni, L.Chenguang, and L.Xinyi,"A survey on intelligent detection for APT attacks,"China Communications, vol. 22, no. 11,
13. R.V.Umasevi and T.R.Nisha Dayana, "A hybrid technique for detecting cyber threats through network traffic analysis," in 2025 5th International Conference on Expert Clouds and Applications (ICOECA), 2025, pp. 1-6.
14. S.Balaba, Y.Chernyshov, A.Skorohodov, and D.Komarov,"Graph-based anomaly detection in industrial control systems," in *2025 IEEE Ural-Siberian Conference on Biomedical Engineering, Radio electronics and Information Technology (USBEREIT)*, 2025, pp. 1-4.
15. S.Aruna, G.L.Prakash, and K.B.Surekha, "Advanced persistent threat detection system using network traffic analysis," in2025International Conference on Electronics and Computing, Communication and Networking Automation Technologies (ICEC2NT), 2025, pp. 1-6.
16. S.Muthumanikandan, G.K.Hegde, P.Swetha, and P.B.Honnavalli, "Flow-based behavioral analysis with machine learning for SSH lateral movement detection,"in 2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA), 2025, pp. 1-5.
17. K.Rathor, V.Keerthika, K.Sunanda, K.Renuga, A.Shobana, and M.Anusuya, "Enhancing network security against APTs through SVM-based network traffic analysis: Identifying anomalies in communication flows," in 2024 International Conference on Computing and Data Science(ICCDS),2024,pp. 1-5.
18. N.H.A.Mutalib,A. Q.M. Sabri, A.W. A.Wahab,E.R.M.F.Abdullah, and N.AlDahoul, "An explainable recursive feature elimination to detect advanced persistent threats using random forest classifier,"in 20253rdInternationalConference on Cyber Resilience (ICCR), 2025, pp. 1-6.
19. Q.Hu, Y.Wang, Z.Su, T.H. Luan, R.Li, and Z.Jiang, "Rethinking online smart contract diagnosis in blockchains: A diffusion perspective,"IEEE Transactions on Networking, vol. 34, pp. 230-245, Sep. 2025.