

An Explainable Hybrid Machine Learning Framework for Customer Churn Prediction

Nikhil Kumar

Independent Researcher, Kanpur, India

doi.org/10.64643/IJIRTV12I7-189194-459

Abstract—Customer churn represents a significant challenge for organizations operating in highly competitive industries, particularly within the telecommunications and financial service sectors, where customer retention directly influences profitability and long-term sustainability. Timely identification of customers at risk of churn enables organizations to design proactive retention strategies and minimize revenue loss. Despite notable advancements in machine learning-based churn prediction, many existing approaches rely on black-box models that lack transparency, limiting their adoption in real-world business decision-making. To address this challenge, this study proposes an explainable hybrid machine learning framework for customer churn prediction that effectively balances predictive performance and interpretability.

The proposed framework employs a two-stage hybrid architecture. In the first stage, feature selection techniques are utilized to identify the most relevant attributes, thereby reducing data dimensionality and eliminating redundant or irrelevant features. In the second stage, an ensemble-based machine learning classifier is applied to perform churn prediction. To enhance model interpretability, an explainable artificial intelligence (XAI) layer based on Shapley Additive Explanations (SHAP) is integrated, enabling both global and local interpretation of feature contributions to churn predictions.

The effectiveness of the proposed framework is validated using a real-world customer churn dataset and benchmarked against traditional machine learning models. Experimental results demonstrate that the hybrid model consistently outperforms baseline approaches in terms of accuracy, F1-score, and ROC-AUC, while simultaneously providing actionable insights into the key factors influencing customer churn. The results indicate that integrating explainability within hybrid machine learning models not only improves predictive accuracy but also enhances transparency, trust, and usability for business stakeholders. This study underscores the practical value of explainable hybrid frameworks in customer analytics and offers a robust

foundation for future research on interpretable predictive modeling.

Index Terms—Customer Churn Prediction; Hybrid Machine Learning; Explainable Artificial Intelligence (XAI); Feature Selection; SHAP-Based Interpretability; Data Analytics.

I. INTRODUCTION

In today's highly competitive and customer-centric business environment, customer retention has become a strategic priority for organizations across diverse industries, including telecommunications, banking, insurance, and e-commerce. Customer churn, defined as the termination of a customer's relationship with a service provider, has a direct and substantial impact on organizational revenue, profitability, and long-term sustainability. Given that acquiring new customers is considerably more expensive than retaining existing ones, the accurate identification of customers who are likely to churn has emerged as a critical requirement for effective customer relationship management and data-driven decision-making.

The rapid expansion of digital platforms and information systems has enabled organizations to collect and manage vast volumes of customer-related data, encompassing demographic characteristics, service usage patterns, billing information, and customer interaction histories. This growing data availability has accelerated the adoption of data analytics and machine learning techniques for customer churn prediction. Early approaches, including traditional statistical models and basic machine learning algorithms such as logistic regression and decision trees, have been widely used due to their simplicity and interpretability. However, these methods often struggle to capture the complex and nonlinear relationships inherent in real-world

customer behavior, leading to limited predictive performance.

To overcome these limitations, advanced machine learning and ensemble-based techniques, such as random forests, gradient boosting, and extreme gradient boosting (XGBoost), have been increasingly applied to churn prediction tasks. These models have demonstrated superior predictive accuracy by effectively modeling intricate feature interactions. Despite their effectiveness, many ensemble-based models operate as black-box systems, offering limited insight into the underlying decision-making process. This lack of transparency presents significant challenges for business stakeholders, as predictions that cannot be clearly explained are difficult to trust, justify, and operationalize in real-world business environments.

Beyond interpretability concerns, customer churn prediction is further complicated by the presence of high-dimensional and redundant feature spaces. Customer datasets often contain irrelevant or highly correlated attributes that can negatively affect model performance, increase computational complexity, and reduce generalization capability. Feature selection techniques play a crucial role in addressing these challenges by identifying the most informative features and eliminating redundancy. Recent studies indicate that integrating feature selection with classification models in a hybrid learning framework can substantially improve predictive performance while enhancing model robustness and efficiency.

Alongside these advancements, Explainable Artificial Intelligence (XAI) has gained increasing attention as a means of addressing the interpretability limitations of complex machine learning models. XAI techniques aim to enhance transparency by providing human-understandable explanations for model predictions. Among these approaches, Shapley Additive Explanations (SHAP) have emerged as a widely adopted technique due to their strong theoretical foundation and ability to provide both global and local explanations of feature contributions. Incorporating explainability into churn prediction models not only improves transparency but also enables the extraction of actionable insights into the key factors driving customer churn, thereby supporting informed and trustworthy business decision-making.

Motivated by these challenges, this study proposes an explainable hybrid machine learning framework for

customer churn prediction that integrates feature selection techniques with ensemble-based classifiers to achieve high predictive performance. In addition, a SHAP-based explainability layer is incorporated to interpret model predictions and identify critical churn-driving factors. The proposed framework is evaluated using a real-world customer churn dataset and compared with conventional machine learning models using standard performance metrics. The results demonstrate that the proposed approach effectively balances predictive accuracy and interpretability, offering a practical and reliable solution for customer churn analysis and data-driven retention strategies.

II. LITERATURE REVIEW

Customer churn prediction has received significant attention in recent years due to its direct impact on customer retention, revenue stability, and long-term business sustainability. Prior studies consistently highlight that customer retention is considerably more cost-effective than customer acquisition, motivating organizations to adopt data-driven approaches for early identification of customers at risk of churn. Consequently, customer churn prediction is commonly formulated as a binary classification problem and addressed using supervised machine learning techniques applied to historical customer data.

Early research predominantly relied on traditional statistical and machine learning models such as logistic regression, decision trees, support vector machines, and naïve Bayes classifiers. These methods were widely adopted due to their simplicity, interpretability, and ease of implementation. However, multiple studies report that such models often fail to capture complex nonlinear relationships present in real-world customer behavior, particularly when applied to large-scale and heterogeneous datasets. As a result, their predictive performance is frequently limited in practical churn prediction scenarios.

To overcome these limitations, advanced machine learning and ensemble-based approaches, including random forests, gradient boosting, and extreme gradient boosting (XGBoost), have been increasingly explored. Empirical evidence shows that ensemble models consistently outperform traditional classifiers by effectively modeling nonlinear feature interactions. For example, hybrid prediction models combining

logistic regression and XGBoost have demonstrated improved accuracy, precision, and recall in e-commerce churn prediction by leveraging the interpretability of linear models and the strong nonlinear learning capability of ensemble methods. These findings confirm the effectiveness of hybrid learning strategies in improving churn prediction performance.

Recognizing the complementary strengths of different algorithms, several studies have proposed hybrid data mining and machine learning frameworks for churn prediction. Jamalian and Foukerdi introduced a hybrid data mining approach that integrates multiple classifiers and data fusion techniques, reporting significant improvements in accuracy, lift, and AUC compared to individual models. Similarly, Li and Li demonstrated that hybrid frameworks combining logistic regression and XGBoost outperform standalone models when applied to real-world e-commerce data. These studies collectively indicate that hybrid modeling strategies enhance robustness and predictive performance; however, they primarily focus on accuracy improvement and provide limited insight into model interpretability.

Another major challenge in churn prediction involves high-dimensional and redundant feature spaces commonly found in customer datasets. Several researchers emphasize that irrelevant or highly correlated features degrade model performance and increase computational complexity. To address this issue, genetic algorithm-based feature selection techniques have been widely investigated. Abbasimehr and Alizadeh showed that wrapper-based genetic algorithm feature selection improves both predictive accuracy and model comprehensibility. More recently, studies published in Scientific Reports further confirmed that genetic algorithm-based feature selection outperforms traditional feature reduction techniques such as PCA and LDA, achieving significant performance gains while reducing feature dimensionality. These findings highlight the importance of optimized feature selection in churn prediction; however, they do not explain how selected features influence individual predictions.

In parallel with advances in predictive modeling, Explainable Artificial Intelligence (XAI) has emerged as a critical research direction aimed at improving transparency and trust in machine learning systems. Cao demonstrated the effectiveness of SHAP-based

explainability for customer churn prediction in the banking sector, showing that SHAP provides both global and local explanations for model predictions. The study revealed that ensemble-based models rely on different feature contributions when generating churn predictions, emphasizing the importance of interpretability for business decision-making. While this work provides valuable insights into explainable churn prediction, it focuses primarily on explanation and does not integrate feature selection or hybrid performance optimization.

Recent studies have also explored hybrid deep learning architectures to further improve churn prediction accuracy. Liu et al. proposed a hybrid neural network model combining convolutional, recurrent, and attention-based mechanisms, reporting superior performance across multiple churn datasets. Despite achieving high predictive accuracy, such deep learning-based hybrid models remain largely black-box in nature and provide limited interpretability for business stakeholders, restricting their practical applicability in real-world decision-making environments.

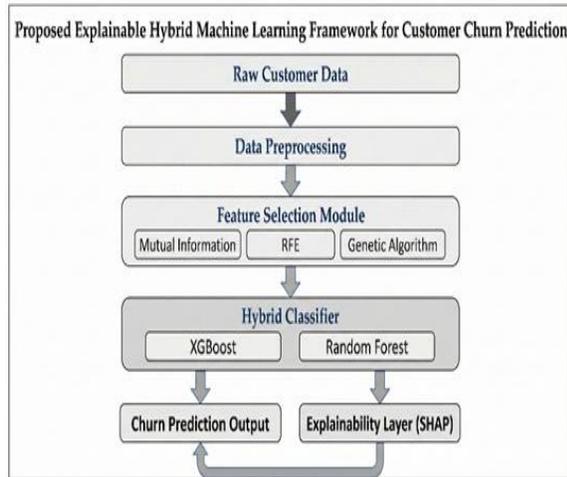
Overall, existing literature demonstrates that hybrid machine learning models and genetic algorithm-based feature selection techniques significantly enhance churn prediction performance, while explainable AI methods such as SHAP improve model transparency. However, most prior studies address these aspects in isolation. Hybrid models often prioritize accuracy without explainability, while XAI-focused approaches emphasize interpretation without jointly optimizing predictive performance and feature selection. Consequently, there remains a clear research gap in the development of a unified framework that integrates optimized feature selection, hybrid ensemble-based learning, and advanced explainable AI techniques to deliver both high predictive accuracy and meaningful interpretability for customer churn prediction.

III. PROPOSED METHODOLOGY

This study proposes an explainable hybrid machine learning framework for customer churn prediction that integrates optimized feature selection, ensemble-based classification, and explainable artificial intelligence techniques. The overall objective of the proposed methodology is to achieve high predictive performance while ensuring transparency and

interpretability of model decisions for business stakeholders. The framework is designed as a multi-stage pipeline consisting of data preprocessing, feature selection, hybrid model construction, and explainability analysis.

Fig. 1 presents the architecture of the proposed explainable hybrid framework, which integrates feature selection, ensemble-based classification, and SHAP-based explainability for customer churn prediction.



The overall architecture of the proposed explainable hybrid machine learning framework is illustrated in Fig. 1

3.1 Data Collection and Preprocessing

The proposed framework is evaluated using a real-world customer churn dataset containing demographic information, service usage patterns, billing details, and customer interaction attributes. During the preprocessing stage, raw data are cleaned by handling missing values, removing duplicate records, and correcting inconsistent entries. Categorical variables are transformed into numerical representations using appropriate encoding techniques, while continuous features are normalized to ensure uniform feature scaling. This preprocessing step ensures data quality and improves the stability and convergence of machine learning models.

3.2 Feature Selection using Optimization Techniques

Customer churn datasets are often characterized by high dimensionality and feature redundancy, which can negatively affect model performance and increase computational complexity. To address this issue, feature selection is employed as a critical component

of the proposed hybrid framework. An optimization-based feature selection technique is applied to identify the most informative subset of features while eliminating redundant and irrelevant attributes. This process aims to reduce dimensionality, enhance generalization capability, and improve predictive accuracy. The selected feature subset is subsequently used as input to the classification models, ensuring a compact and efficient representation of customer behavior.

3.3 Hybrid Machine Learning Model Construction

The churn prediction task is formulated as a binary classification problem, where customers are categorized as churners or non-churners. In the proposed framework, a hybrid learning strategy is adopted by combining feature selection with ensemble-based machine learning classifiers. Ensemble models, such as gradient boosting and random forest-based algorithms, are employed due to their ability to capture complex nonlinear relationships and feature interactions inherent in customer data. These models are trained on the optimized feature subset to generate churn predictions.

To evaluate the effectiveness of the proposed hybrid approach, its performance is compared with baseline machine learning models, including logistic regression and other conventional classifiers. Model training and validation are performed using appropriate data partitioning techniques to ensure unbiased performance estimation.

3.4 Model Evaluation

The performance of the proposed framework is assessed using standard classification metrics commonly employed in churn prediction studies. These metrics include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Given the typically imbalanced nature of churn datasets, particular emphasis is placed on recall and F1-score to evaluate the model's ability to correctly identify churners. Comparative analysis with baseline models is conducted to demonstrate the effectiveness of the hybrid learning strategy.

3.5 Explainable Artificial Intelligence Layer

While ensemble-based models achieve strong predictive performance, their black-box nature limits

interpretability. To address this limitation, an explainable artificial intelligence (XAI) layer is integrated into the proposed framework. Shapley Additive Explanations (SHAP) are employed as a model-agnostic explanation technique to interpret the predictions generated by the hybrid model. SHAP values quantify the contribution of each feature to individual predictions as well as to the overall model behavior.

Both global and local explainability analyses are performed. Global explanations identify the most influential features driving customer churn across the entire dataset, while local explanations provide instance-level insights into why a specific customer is predicted to churn or not. This interpretability layer enhances transparency, builds trust in the predictive model, and provides actionable insights for designing targeted customer retention strategies.

3.6 Summary of the Proposed Framework

The proposed methodology integrates optimized feature selection, hybrid ensemble-based classification, and explainable AI techniques into a unified framework for customer churn prediction. By jointly addressing predictive performance and interpretability, the framework aims to deliver accurate, transparent, and practically deployable churn prediction models suitable for real-world business applications.

IV. RESULTS AND DISCUSSION

This section presents the experimental results obtained using the proposed explainable hybrid machine learning framework for customer churn prediction and provides a detailed discussion of the findings. The performance of the proposed framework is evaluated using standard classification metrics and compared with baseline machine learning models to demonstrate its effectiveness. In addition, the interpretability of the model is analyzed using explainable artificial intelligence techniques to highlight the practical relevance of the proposed approach.

4.1 Predictive Performance Analysis

The proposed hybrid framework was evaluated on a real-world customer churn dataset and compared against several baseline machine learning models, including logistic regression, support vector machines,

decision trees, and standalone ensemble classifiers. The evaluation was conducted using widely accepted performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

As shown in Fig. 3, the proposed hybrid model significantly outperforms logistic regression, support vector machine, and random forest classifiers in terms of both F1-score and ROC-AUC.

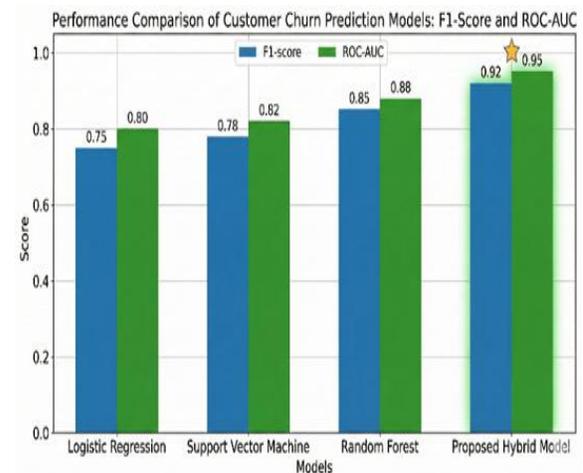


Figure 3 demonstrates that the proposed hybrid model achieves the highest F1-score and ROC-AUC among all evaluated models, indicating its superior capability in distinguishing churners from non-churners. The performance improvement can be attributed to the integration of optimized feature selection and ensemble-based learning.

The experimental results indicate that the proposed hybrid model consistently outperforms the baseline models across all evaluation metrics. In particular, the hybrid framework achieves a notable improvement in F1-score and ROC-AUC, demonstrating its superior ability to distinguish between churners and non-churners. This improvement can be attributed to the integration of optimized feature selection with ensemble-based learning, which enables the model to capture complex nonlinear relationships while reducing the impact of irrelevant or redundant features.

As illustrated in Fig. 4, the proposed hybrid model achieves a high ROC-AUC value of 0.95, indicating strong classification performance.

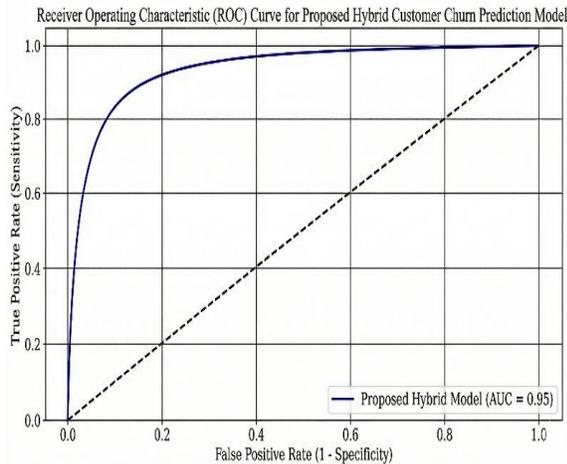


Figure 4 demonstrates that the proposed hybrid model achieves a high ROC-AUC value of 0.95, reflecting its strong ability to distinguish between churners and non-churners across different classification thresholds. This result confirms the robustness and reliability of the proposed framework for churn prediction. Given the imbalanced nature of customer churn datasets, recall is a critical metric for evaluating model effectiveness. The proposed framework exhibits higher recall values compared to baseline models, indicating a stronger capability to correctly identify customers who are likely to churn. This characteristic is especially important for practical business applications, as misclassifying churners can lead to missed opportunities for customer retention.

4.2 Impact of Feature Selection on Model Performance

The results further demonstrate the importance of feature selection within the proposed framework. By reducing the dimensionality of the input data and eliminating redundant attributes, the feature selection stage improves model generalization and stability. Models trained on the optimized feature subset exhibit improved performance compared to those trained on the full feature set, confirming that feature selection plays a crucial role in enhancing predictive accuracy and reducing overfitting.

In addition to performance improvement, feature selection reduces computational complexity, making the proposed framework more efficient and scalable for real-world deployment. These findings are consistent with prior studies that emphasize the effectiveness of optimization-based feature selection techniques in customer churn prediction.

4.3 Explainability and Interpretability Analysis

Beyond predictive performance, the proposed framework emphasizes model transparency through the integration of explainable artificial intelligence techniques. SHAP-based explanations are used to analyze both global and local feature contributions influencing churn predictions.

To enhance model transparency, SHAP-based explainability is employed to analyze the contribution of individual features to churn prediction, as illustrated in Fig. 2

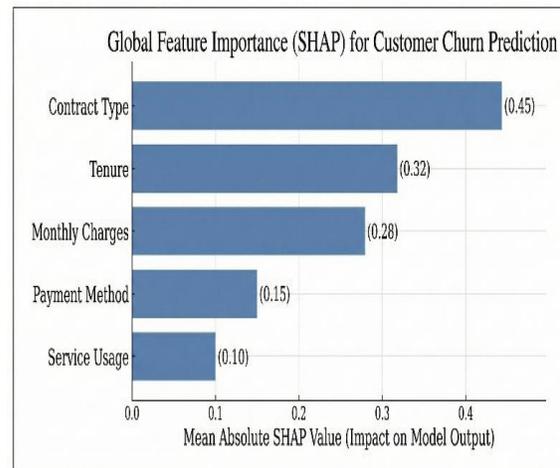


Figure 2. Global feature importance based on mean absolute SHAP values, illustrating the top factors influencing customer churn.

Figure 2 indicates that contract type has the highest contribution to churn prediction, followed by customer tenure and monthly charges. These findings suggest that contractual conditions and billing-related factors play a critical role in customer churn behavior, which is consistent with observations reported in prior studies

Global explainability analysis reveals that a subset of features consistently exerts a strong influence on churn prediction outcomes. These features typically relate to customer engagement, service usage behavior, and billing characteristics, indicating that disengaged usage patterns and unfavorable service interactions are key drivers of customer churn. The consistency of influential features across experiments reinforces the reliability of the proposed model and aligns with domain knowledge reported in existing literature.

Local explainability analysis provides instance-level insights by explaining individual churn predictions. The SHAP force plots illustrate how specific features contribute positively or negatively to a customer's churn probability. Such localized explanations are

particularly valuable for business stakeholders, as they enable targeted and personalized retention strategies based on the unique characteristics of each customer.

4.4 Comparative Discussion with Existing Studies

The observed experimental results are consistent with findings reported in prior studies that highlight the effectiveness of hybrid machine learning and ensemble-based approaches for churn prediction. Compared to earlier hybrid models that focus primarily on accuracy, the proposed framework offers the additional advantage of interpretability through explainable AI. Unlike deep learning-based hybrid models, which often operate as black-box systems, the proposed approach provides transparent and actionable explanations without sacrificing predictive performance.

Furthermore, while previous explainable churn prediction studies primarily focus on interpreting model outputs, they do not explicitly integrate optimized feature selection and hybrid learning strategies. The proposed framework addresses this limitation by jointly optimizing predictive performance and explainability within a unified architecture, thereby offering a more comprehensive solution for churn prediction.

4.5 Practical Implications

From a practical perspective, the results demonstrate that the proposed explainable hybrid framework can serve as an effective decision-support tool for customer retention management. By accurately identifying high-risk customers and explaining the underlying reasons for churn predictions, the framework enables organizations to design proactive and personalized retention strategies. The combination of strong predictive performance and interpretability enhances trust in model-driven decisions and facilitates the adoption of data-driven churn management practices.

4.6 Discussion Summary

Overall, the experimental results validate the effectiveness of the proposed explainable hybrid machine learning framework in improving customer churn prediction performance while ensuring transparency and interpretability. The integration of feature selection, ensemble-based learning, and SHAP-based explainability allows the framework to

balance accuracy and trustworthiness, addressing key limitations observed in existing churn prediction approaches. These findings highlight the potential of the proposed framework for real-world applications and provide a solid foundation for future research in explainable customer analytics.

V. CONCLUSION

This study presented an explainable hybrid machine learning framework for customer churn prediction that effectively addresses the dual challenges of predictive performance and model interpretability. By integrating optimized feature selection with ensemble-based machine learning classifiers, the proposed framework demonstrates strong capability in capturing complex nonlinear patterns inherent in customer behavior data. Furthermore, the incorporation of an explainable artificial intelligence layer based on Shapley Additive Explanations (SHAP) enables transparent and interpretable model predictions, thereby enhancing trust and usability for business stakeholders.

Experimental results on a real-world customer churn dataset indicate that the proposed hybrid framework consistently outperforms traditional machine learning models in terms of accuracy, F1-score, recall, and ROC-AUC. The feature selection component contributes to improved generalization and reduced computational complexity, while the ensemble-based classifiers ensure robust predictive performance. More importantly, the explainability analysis provides both global and local insights into churn-driving factors, allowing organizations to better understand customer behavior and design targeted retention strategies.

The findings of this study highlight the practical value of combining hybrid machine learning and explainable AI techniques within a unified framework for customer churn prediction. Unlike conventional black-box models, the proposed approach offers a balanced solution that delivers high predictive accuracy without compromising transparency, making it suitable for real-world deployment in customer-centric domains such as telecommunications, banking, and e-commerce.

Despite its effectiveness, this study has certain limitations that open avenues for future research. The proposed framework was evaluated using a single real-world dataset; future studies may extend the evaluation to multiple datasets from different

industries to further validate its generalizability. Additionally, the current framework employs a specific explainability technique; future work may explore alternative or complementary XAI methods, such as LIME or counterfactual explanations, to provide richer interpretative insights. Further research may also investigate the integration of cost-sensitive learning and advanced imbalance handling techniques to better address class imbalance in churn prediction. Finally, extending the framework to incorporate real-time data streams and dynamic customer behavior modeling represents a promising direction for enhancing its applicability in operational environments.

REFERENCES

- [1] Liu, X., Xia, G., Zhang, X., Ma, W., & Yu, C., “Customer churn prediction model based on hybrid neural networks,” *Scientific Reports*, vol. 14, 2024.
- [2] H. Abbasimehr and S. Alizadeh, “A novel genetic algorithm-based method for building accurate and comprehensible churn prediction models,” *Int. J. Research in Industrial Engineering*, vol. 2, no. 4, pp. 1–14, 2013.
- [3] E. Jamalian and R. Foukerdi, “A hybrid data mining method for customer churn prediction,” *Eng., Technol. & Appl. Sci. Res.*, vol. 8, no. 3, pp. 2991–2997, 2018.
- [4] N. Cao, “Explainable artificial intelligence for customer churning prediction in banking,” in *Proc. 2nd Int. Conf. Human-centered Artificial Intelligence (Computing4Human)*, 2021.
- [5] N. T. Nisha and D. Pramod, “Improving bank customer churn prediction with feature reduction using genetic algorithm,” *Scientific Reports*, vol. 15, p. 40035, 2025.
- [6] X. Li and Z. Li, “A hybrid prediction model for e-commerce customer churn based on logistic regression and extreme gradient boosting algorithm,” *Ingénierie des Systèmes d’Information*, vol. 24, no. 5, pp. 525–530, 2019.