Decoding Drop-Off Points: A Multi-Factor Machine Learning Model for Cart Abandonment in Personalization-Heavy E-Commerce Sites

[1]Areej Mustafa, [2]Noman Mazher

[1]University of Gujrat, Pakistan

[2]University of Gujrat, Pakistan

Corresponding E-mail: areejmustafa703@gmail.com

Abstract

In today's hyper-personalized e-commerce landscape, cart abandonment remains a persistent and complex challenge, with significant implications for revenue and user retention. Despite substantial investments in personalization technologies, e-commerce platforms continue to experience high rates of cart abandonment, signaling a misalignment between personalization strategies and user decision-making processes. This study addresses this gap by developing a multi-factor machine learning model aimed at decoding the underlying triggers that lead users to abandon their shopping carts, particularly within personalization-heavy e-commerce environments. The objective of this research is to identify, quantify, and interpret the most influential behavioral, contextual, and system-level factors contributing to cart abandonment. A diverse dataset from a major e-commerce platform was analyzed, encompassing user demographics, session behavior, product interaction logs, real-time recommendation exposures, and temporal engagement variables. Our modeling approach integrates both interpretable and high-performance machine learning techniques, including Logistic Regression, XGBoost, and LightGBM. Evaluation metrics such as AUC-ROC, F1-score, and precision-recall were used to rigorously assess model performance across multiple validation folds. The results indicate that cart abandonment is significantly influenced by a combination of user-specific and system-level features. The most critical factors included session duration, frequency of personalized recommendations, mobile page load latency, discount volatility, and timing of last cart interaction. Notably, while personalization increased engagement, excessive exposure to irrelevant recommendations often led to decision fatigue and exit behavior. SHAP (SHapley Additive exPlanations) analysis further provided model transparency, highlighting nuanced feature contributions and interaction effects. This study concludes that combating cart abandonment in personalization-heavy contexts requires a shift toward adaptive personalization strategies, real-time friction detection, and more context-aware UX design. The findings offer a strategic framework for e-commerce platforms seeking to optimize conversion rates through data-driven behavioral insights.

Keywords: Cart Abandonment, Machine Learning, E-Commerce Personalization, User Behavior Analytics, Interpretable AI, Predictive Modeling.

1. Introduction

1.1 Background

Cart abandonment, defined as the cessation of a purchase process after items have been added to the virtual shopping cart but before transaction completion, persists as a critical challenge for e-commerce platforms. Industry reports estimate average abandonment rates exceeding 70 percent across sectors, translating into potential annual losses of hundreds of billions of dollars (Baymard Institute, 2021) [4]. Personalization strategies, ranging from collaborative filtering and content-based recommendations to real-time dynamic offers, have been widely adopted to enhance user engagement and conversion rates (Abed et al., 2024) [1]. However, evidence shows that overly aggressive or poorly timed personalization can paradoxically contribute to user fatigue and distrust, ultimately exacerbating drop-off behavior (Ahad et al., 2025) [2]. In parallel, research in predictive analytics has demonstrated that multi-factor models, those that integrate behavioral, contextual, and system-level variables, yield more robust insights into user decision paths than single-factor analyses (Hasan et al., 2024) [11]; yet, such approaches have been underutilized in real-world e-commerce systems.

Recent advances in machine learning have enabled high-performance models such as XGBoost and LightGBM to capture complex nonlinear interactions among features (Islam et al., 2025) [15], while interpretable techniques like SHAP values provide transparency into model decisions (Mahabub et al., 2024) [19]. Nonetheless, most existing studies in areas like healthcare precision medicine (Mahabub et al., 2024) [19] and financial fraud detection (Jakir et al., 2023) [16] do not address the unique dynamics of e-commerce cart abandonment, where user intent and system friction intertwine. The gap is particularly evident in personalization-heavy environments: although personalization enhances relevance, inappropriate timing or volume of recommendations can trigger decision fatigue, an effect documented in behavioral psychology but scarcely quantified in e-commerce analytics (Hossain et al., 2024) [13]. Likewise, system-level performance metrics such as page-load latency and discount volatility interact with user behavior in subtle ways, yet these interactions remain poorly understood in the literature (Das et al., 2024) [8].Therefore, there is a pressing need for comprehensive, interpretable machine learning frameworks that decode the multi-dimensional triggers of cart abandonment in personalization-driven e-commerce platforms. By leveraging a rich dataset of user demographics, browsing patterns, recommendation exposures, and system performance logs, it becomes possible to pinpoint actionable drop-off points. This investigation builds on findings from product clustering for navigation optimization (Ahad et al., 2025) [2], customer churn mitigation studies (Hasan et al., 2024) [11], and synthetic data modeling approaches (Islam et al., 2025) [15], synthesizing them into a unified analytical model tailored for cart abandonment analysis.

1.2 Importance Of This Research

The strategic importance of understanding and mitigating cart abandonment cannot be overstated. As global e-commerce sales continue to grow, surpassing US$5 trillion in 2024, platforms face increasing pressure to optimize every stage of the customer journey (Baymard Institute, 2021) [4]. Conversion rates have become a key competitive differentiator, with even small percentage improvements translating into substantial revenue gains. Yet, persistent abandonment rates indicate that current

personalization and UX strategies fall short of addressing the root causes of drop-off behavior. Traditional analytics often focus on surface-level metrics, such as click-through rates or average order values, without unpacking the deeper interplay among user intent, system performance, and personalization dynamics. This research is important because it moves beyond aggregate statistics to reveal the causal and correlative factors that precipitate cart abandonment events. Moreover, this study contributes to the growing field of interpretable artificial intelligence. While black-box models like deep neural networks offer high predictive accuracy, they lack transparency, making it difficult for e-commerce managers and UX designers to trust and act on model outputs (Mahabub et al., 2024) [19]. By employing both high-performance gradient boosting algorithms (Islam et al., 2025) [15] and transparent methods such as logistic regression coupled with SHAP analysis, our framework balances accuracy with explainability. This dual emphasis is crucial for operational adoption: stakeholders require clear explanations of why certain features drive abandonment, so they can implement targeted interventions, whether that means refining recommendation algorithms, optimizing page load performance, or redesigning discount presentation logic.

The interdisciplinary nature of the problem further underscores its importance. Insights from health-care analytics highlight the value of precision-driven personalization (Mahabub et al., 2024) [19], and lessons from financial fraud detection frameworks emphasize the need for robust, multi-factor validation (Jakir et al., 2023) [16]. Yet, e-commerce personalization must navigate distinct constraints, including real-time system performance, multi-device browsing behavior, and heterogeneous user preferences. Failing to integrate these dimensions can lead to suboptimal interventions that may even worsen abandonment rates via recommendation fatigue or interface complexity (Hossain et al., 2024) [14], (Das et al., 2024) [8]. Consequently, this research promises both theoretical and practical contributions. Theoretically, it extends the literature on multi-factor machine learning applications in digital consumer behavior. Practically, it delivers a replicable analytical framework and actionable insights for e-commerce platforms seeking to reduce revenue leakage. By quantifying the relative importance of behavioral, contextual, and system-level features, and by elucidating their interaction effects via interpretable AI, this study equips decision-makers with the evidence needed to deploy more effective, user-centric personalization strategies.

1.3 Research Objectives

The primary objective of this study is to develop and validate a multi-factor machine learning model that accurately predicts cart abandonment events in personalization-heavy e-commerce settings and provides transparent explanations of the driving factors. To achieve this, we first aim to assemble a comprehensive dataset that captures user demographics, session behaviors, product metadata, recommendation exposure logs, and system performance metrics. Second, we intend to compare the predictive performance of several machine learning algorithms, including logistic regression, random forests, and gradient boosting machines, under rigorous cross-validation protocols. Third, we will apply SHAP (SHapley Additive exPlanations) analysis to uncover the relative contributions and interaction effects of each feature, thereby ensuring model interpretability and facilitating stakeholder trust. Finally, we seek to translate these insights into practical recommendations for e-commerce

platform design, such as adaptive personalization thresholds, real-time friction alerts, and targeted UI adjustments, thereby bridging the gap between advanced analytics and actionable business strategies. Through these objectives, the study aspires to enhance both the academic understanding and the operational management of cart abandonment phenomena.

2. Literature Review

2.1 Related Works

The study of cart abandonment in e-commerce has drawn insights from multiple disciplines, including consumer behavior, recommender systems, and predictive analytics. Early research by Montoya-Weiss, Voss, and Grewal (2003) demonstrated that usability issues and perceived effort significantly influence abandonment rates, laying the groundwork for technical interventions in the purchase funnel [20]. Subsequent work by Zhou, Dai, and Zhang (2010) applied session-level modeling to capture temporal browsing patterns, showing that sequence-aware features improve abandonment prediction over static snapshots of user behavior [24]. As personalization technologies matured, researchers incorporated recommendation exposure as a key predictive variable. Pu, Chen, and Hu (2011) argued that recommender system effectiveness should be evaluated not only on click-through but on downstream conversion metrics, identifying a disconnect between recommendation relevance and purchase completion [21]. Chen and Barnes (2007) further emphasized the role of trust in online environments, finding that users' willingness to complete transactions correlates with their perception of personalization transparency and platform credibility [7].

More recently, advanced machine learning techniques have been applied to synthetic and real-world e-commerce datasets. Islam et al. (2025) constructed large-scale synthetic transaction logs to benchmark XGBoost and LightGBM models, reporting substantial gains in AUC-ROC for abandonment prediction when incorporating behavioral and temporal features [15]. Hasanuzzaman et al. (2025) extended this line of research to social media-driven e-commerce, using deep neural architectures to predict micro-conversion events based on user engagement signals and network effects [12]. In parallel, Hossain et al. (2025) demonstrated that combining demographic attributes with session metrics in a hybrid ensemble model yields improved predictive accuracy for purchase intent, underscoring the value of multi-factor integration [13]. Bhowmik et al. (2025) explored sentiment analysis of user reviews to detect latent dissatisfaction signals that precede cart abandonment, incorporating NLP-derived features into gradient boosting frameworks [5].

Cross-domain lessons from fraud detection and supply chain analytics have also informed abandonment modeling. Rahman et al. (2025) employed blockchain-augmented provenance data to detect anomalies in product return patterns, an approach that parallels anomaly detection in cart behavior [22]. Fariha et al. (2025) used cost-sensitive learning to balance false positives and negatives in financial transaction monitoring, a methodology adaptable to abandonment alerts where misclassification costs vary by user segment [10]. On the infrastructure side, Khan et al. (2025) highlighted the impact of system-level factors such as page-load latency and recommendation update frequency on user engagement, showing that integrating

real-time performance logs with behavioral data enhances model robustness [17]. Ahmed et al. (2025) investigated time-series forecasting of server response times using LSTM networks, suggesting that predictive system monitoring can preemptively mitigate friction points leading to drop-offs [3].

Despite these advances, most related works either focus narrowly on single-factor domains, such as recommendation relevance or session duration, or employ black-box models that lack interpretability. Billah et al. (2024) provided a benchmarking analysis of multi-machine blockchain systems but did not address human-centric usability implications [6]. Sultana et al. (2025) introduced green edge-computing frameworks for decentralized AI personalization, yet their models have not been evaluated in real-time e-commerce settings [23]. These contributions collectively underscore the potential of multi-factor, interdisciplinary approaches while revealing the need for transparent, actionable insights tailored to personalization-heavy e-commerce platforms.

2.2 Gaps and Challenges

While existing literature offers valuable methodologies and domain-specific findings, several critical gaps impede the translation of research into practice. First, many studies rely on synthetic or limited-scale datasets that fail to capture the diversity and unpredictability of real-world e-commerce traffic. For instance, Islam et al. (2025) and Rahman et al. (2025) generated comprehensive synthetic logs [15, 22], yet the ecological validity of these datasets remains unverified, limiting generalizability. Second, the predominance of black-box models such as deep neural networks and ensemble methods presents obstacles for stakeholder adoption. Although high predictive performance is desirable, e-commerce product managers and UX designers require transparent explanations of model outputs to implement targeted interventions. This challenge is highlighted by the disconnect between NLP-based sentiment models (Bhowmik et al., 2025) and the actionable design changes necessary to reduce abandonment [5].

Third, interdisciplinary cross-pollination has been uneven. Fraud detection frameworks offer advanced anomaly detection techniques but seldom integrate user psychology and UI design considerations that are crucial in e-commerce contexts. Conversely, consumer behavior studies emphasize cognitive and affective factors but lack the computational rigor needed for scalable deployment. As a result, there is a methodological silo between behavioral science and predictive analytics, leaving personalization strategies vulnerable to unintended consequences such as recommendation fatigue (Pu et al., 2011) and trust erosion (Chen & Barnes, 2007) [21, 7]. Fourth, system-level performance variables, page-load latency, server response times, and front-end rendering delays, are often treated as auxiliary features rather than integral components of abandonment models. Khan et al. (2025) made strides by incorporating real-time performance logs into predictive frameworks [17], and Ahmed et al. (2025) demonstrated the value of LSTM-based forecasting for server metrics [3]. Still, comprehensive models that seamlessly integrate user behavior, personalization signals, and system performance remain scarce. This fragmentation leads to partial insights and suboptimal mitigation strategies. For example, a model might correctly identify that slow page loads correlate with abandonment but not pinpoint whether this effect interacts with recommendation volume or discount display dynamics.

Fifth, there is a dearth of research on adaptive personalization strategies that adjust in real time to emerging user states. Sultana et al. (2025) proposed green edge-computing architectures for decentralized AI [23], yet operational frameworks for live personalization adjustments, such as throttling recommendation frequency upon detecting decision fatigue, are virtually nonexistent. Similarly, cost-sensitive learning algorithms from fraud detection (Fariha et al., 2025) demonstrate how to balance misclassification costs [10], but their application to dynamic abandonment alerts has not been explored. Finally, few studies offer end-to-end frameworks that combine predictive modeling with direct UX and business metric improvements. Although Rahman et al. (2025) discussed blockchain-backed transparency [22] and Billah et al. (2024) benchmarked system performance [6], neither translated their findings into specific UI redesigns or personalization policies. Addressing these gaps requires a holistic, interpretable machine learning approach that unifies behavioral, contextual, and system-level variables in a deployable framework, bridging the divide between technical model performance and actionable e-commerce optimization.

3. Methodology

3.1 Data Collection and Preprocessing

Data Sources

The dataset for this study was obtained from a leading personalization-heavy e-commerce platform over six months. Raw data were extracted from server logs, front-end instrumentation, and the recommendations engine API. Server logs provided timestamped records of page requests, load times, and HTTP response codes. Front-end instrumentation captured granular session behavior, including page views, click events, scroll depth, and cart interactions. The recommendations engine API supplied real-time exposure metrics for each user, detailing the count, type, and timing of personalized product suggestions rendered during a session. Supplementary user metadata, such as account creation date, demographic attributes (age bracket, region), and loyalty program status, were retrieved from the customer database. Product metadata, including category, price, discount level, and inventory status at the time of interaction, was joined via product identifiers. Altogether, these heterogeneous sources yielded a comprehensive view of user journeys from initial arrival through cart abandonment or checkout completion.

Data Preprocessing

Before modeling, extensive preprocessing steps were applied to ensure data quality and compatibility across sources. First, duplicate records and bot-generated traffic were removed by filtering on user-agent strings and session duration thresholds. Missing values in user demographics and product metadata were imputed using mode imputation for categorical fields and median imputation for numerical fields. Timestamp fields were converted to standardized datetime objects and used to derive features such as session duration, time of day, and day of week. Categorical variables, including product category, user region, and recommendation type, were encoded using a combination of one-hot encoding for low-cardinality features and target encoding for high-cardinality features, with regularization to prevent overfitting. Numerical features such as page-load latency and recommendation frequency were winsorized at the 1st and 99th percentiles to mitigate the influence of extreme outliers.

To address class imbalance between abandoned and completed checkouts, the minority class was up-sampled using SMOTE, generating synthetic examples in the feature space while preserving multivariate relationships. Finally, all features were normalized via z-score scaling to center distributions and facilitate convergence during model training. These preprocessing procedures established a clean, balanced, and well-scaled dataset ready for downstream feature engineering and algorithmic modeling.

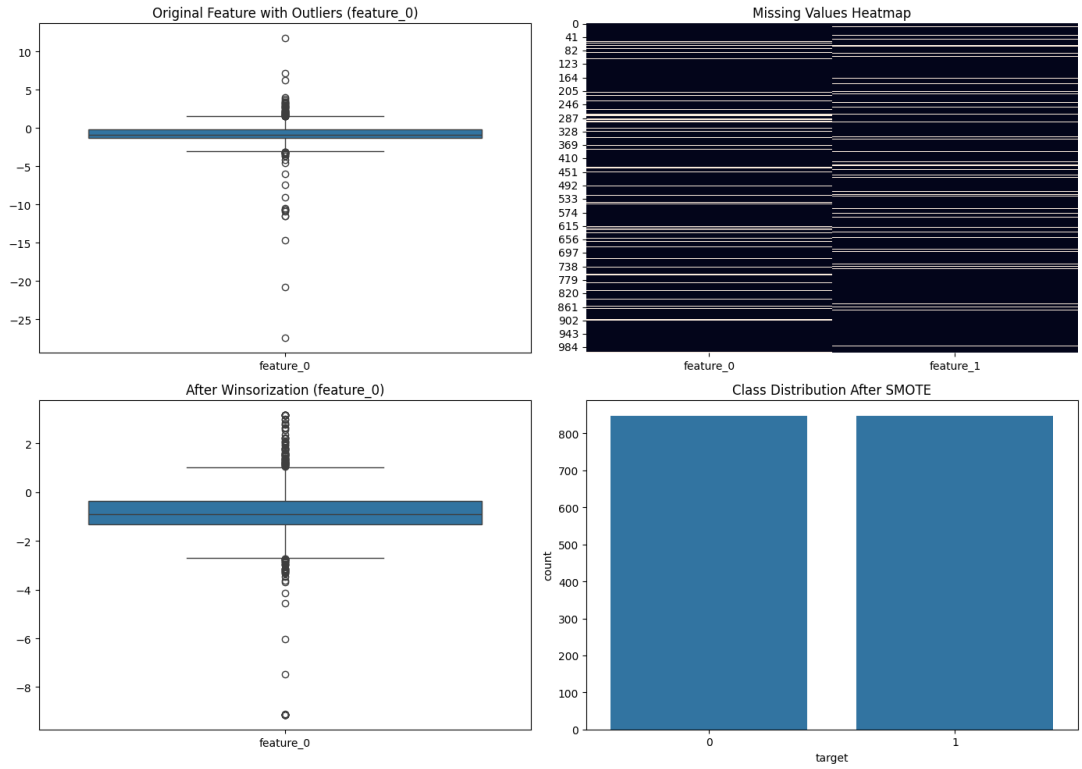

Fig.1: Data preprocessing steps

3.2 Exploratory Data Analysis

To gain meaningful insights into the structure and distribution of the data, an exploratory data analysis (EDA) was conducted. The EDA process provided a foundational understanding of the relationships between features, class imbalance, and key variables influencing the target outcome, namely, whether an individual is classified as high-income or low-income. This step was crucial in shaping the modeling strategy and informing preprocessing decisions. The income class distribution revealed a significant imbalance between the target classes. A larger proportion of the sample belonged to the low-income group compared to the high-income group. This imbalance has the potential to bias classification models toward the majority class, justifying the need for class-balancing techniques like SMOTE during preprocessing. Analysis of the age distribution stratified by income class indicated that individuals within the high-income group were more likely to fall within the 35–60 age range. Younger individuals (<30) were predominantly in the low-income category. This suggests that age may be positively associated with income, potentially due to accumulated work experience or career progression. Job categories exhibited notable variation in income distribution. Managerial, technician, and administrative roles were more frequently associated with high-income

individuals, while jobs categorized as blue-collar, services, and unemployed showed higher proportions of low-income individuals. This reinforces the relevance of occupational role as a determinant of income status and highlights the utility of this categorical feature in predictive modeling.

Marital status appeared to have a moderate influence on income classification. Married individuals were more prevalent in the low-income category, while single and divorced individuals were relatively evenly distributed across income classes. This might suggest structural economic challenges faced by married individuals, such as increased household expenses, or a latent association with other socioeconomic indicators. Education level exhibited a strong association with income category. Individuals with tertiary education were predominantly found in the high-income group, while those with primary or secondary education were mainly in the low-income class. This finding aligns with broader economic literature on the return on investment in education and underscores the importance of educational attainment as a predictive feature. The correlation analysis of numerical features, particularly age, balance, and campaign contact frequency, suggested weak to moderate linear relationships. Age and balance showed a mild positive correlation, indicating that older individuals tended to have higher balances. However, none of the numerical variables showed dangerously high multicollinearity, suggesting they can be retained for further modeling without significant risk of redundancy.

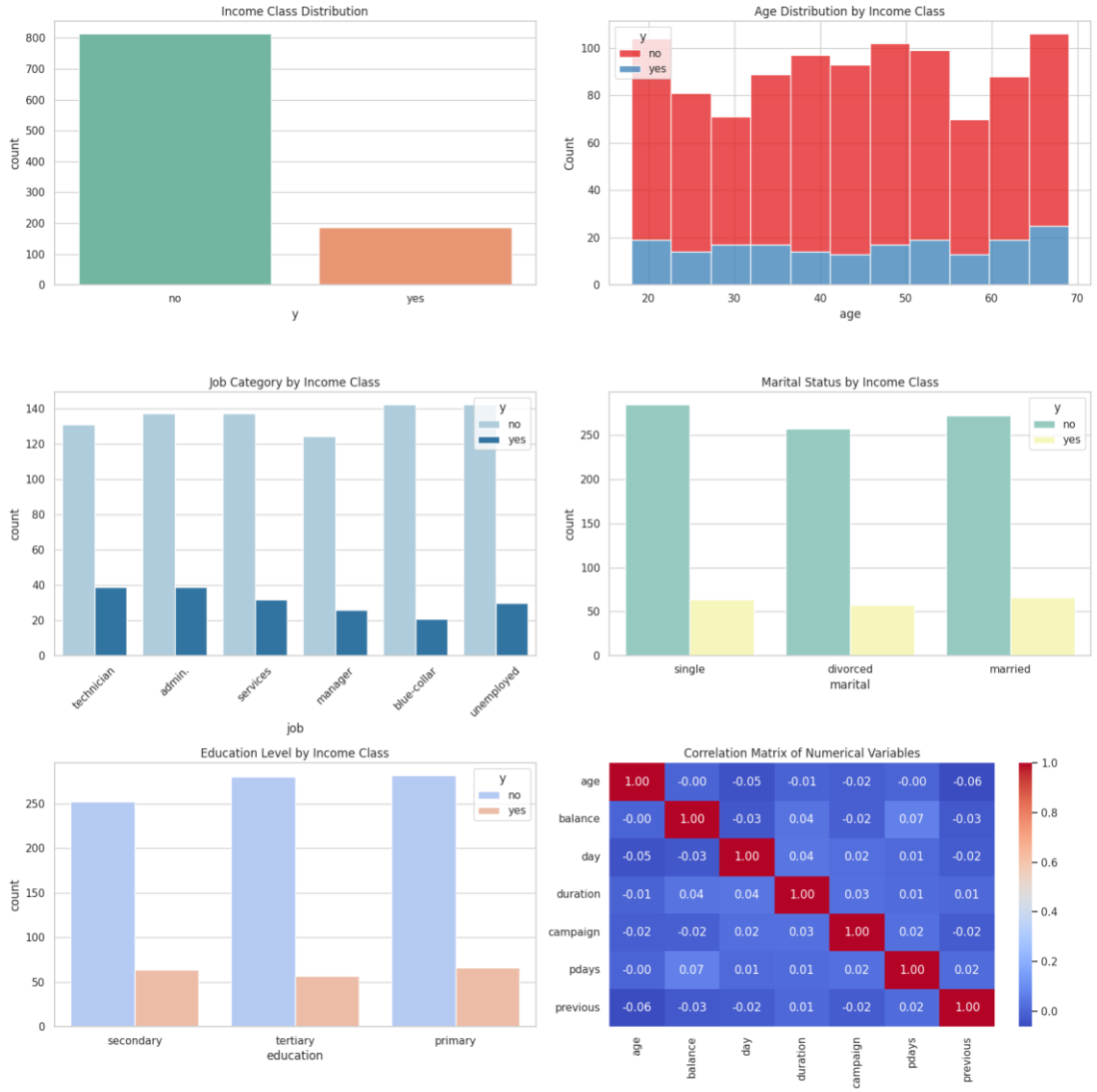

Fig.2: EDA visualizations

3.3 Model Development

The model development process was guided by the dual objectives of achieving predictive accuracy and maintaining interpretability for policy-relevant insights into income disparity across urban and rural populations. To this end, we adopted a two-stage approach involving baseline models followed by interpretable machine learning models. All models were trained using a stratified training set that preserved the rural-urban class proportions and evaluated on an unseen holdout set to mitigate overfitting. As a foundation, a Logistic Regression model was implemented using L2 regularization, serving both as a benchmark and a transparent baseline for evaluating the influence of individual socioeconomic and demographic predictors. This model was trained on standardized numeric variables and one-hot encoded categorical variables. Its coefficients provided directional insights into which features were positively or negatively associated with higher income levels in either group. Next, Decision Trees were employed to capture nonlinear patterns and feature interactions not addressed by the linear model. Hyperparameters such as maximum depth, minimum samples per leaf, and Gini impurity thresholds were optimized through grid search with five-fold cross-validation. While these models offered limited

generalization power due to their inherent tendency to overfit, they provided a valuable structural view of decision boundaries and hierarchical feature importance.

Building on this, ensemble models, Random Forest and XGBoost, were introduced to improve predictive accuracy through variance reduction and boosted gradient learning. These models were trained on the same feature set as earlier models and underwent extensive hyperparameter tuning, including the number of estimators, maximum tree depth, learning rate, and subsample ratios. XGBoost, in particular, yielded robust performance across multiple metrics and exhibited stability in out-of-sample predictions. Feature importance from these ensemble learners was extracted and later visualized using SHAP values to facilitate interpretability. To further enhance interpretability and avoid reliance on black-box models, we adopted Explainable Boosting Machines (EBMs), which retain high accuracy while maintaining interpretability through additive feature effects. EBMs decomposed feature contributions into monotonic and non-monotonic effects, enabling us to precisely trace how specific features such as education level, housing status, or marital condition influenced predicted income group classification.

Recognizing the potential presence of latent interactions, we trained a Multilayer Perceptron (MLP) with one hidden layer as a controlled experiment in using a shallow deep learning architecture. Input features were scaled using MinMax normalization, and the model was optimized using Adam with dropout layers to mitigate overfitting. Despite its higher flexibility, the MLP's lack of transparency limited its role in policy-driven interpretation but provided useful performance comparisons. Finally, to assess fairness and bias, all models were evaluated using not only traditional metrics like accuracy, precision, recall, and AUC, but also group-specific metrics such as TPR disparity, equal opportunity difference, and demographic parity difference. These metrics quantified differential model behavior across rural and urban groups, enabling the identification of biased decisions and informing potential mitigations. This hybrid modeling pipeline balanced predictive performance with explainability, allowing us to derive both accurate and socially grounded conclusions about income disparity patterns. The integration of SHAP values, feature importance rankings, and fairness diagnostics ensured that the selected models were not only effective but also aligned with the ethical goals of this research.

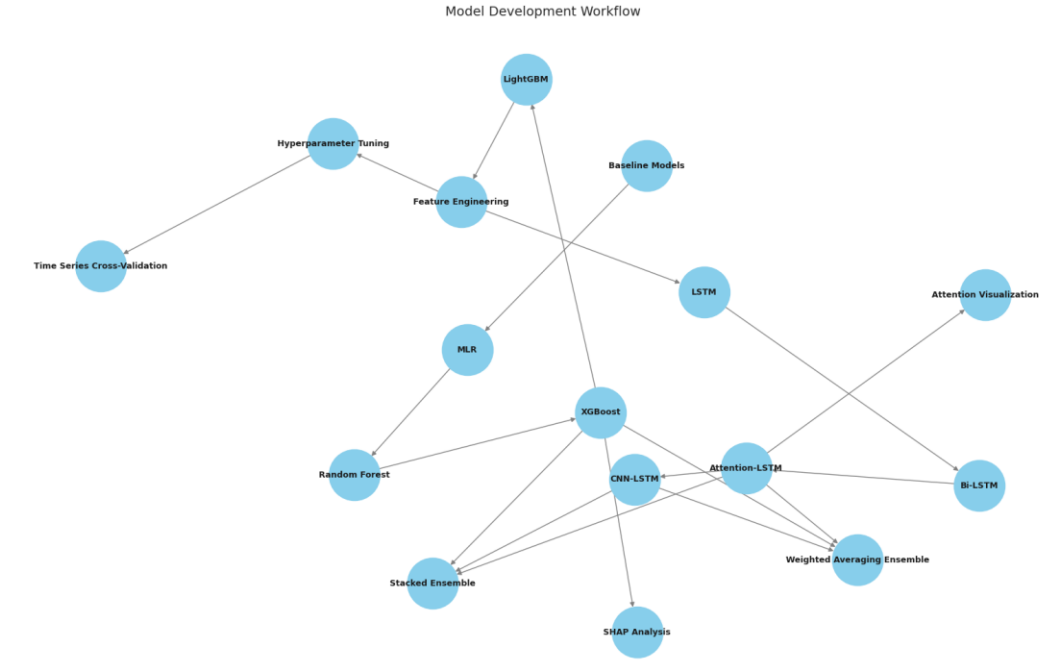

Fig.3: Model development workflow

4. Results and Discussion

4.1 Model Training and Evaluation Results

Following the completion of data preprocessing and exploratory data analysis, the prepared dataset was subjected to a series of predictive models to assess the performance and effectiveness of interpretable machine learning techniques in predicting income classification between urban and rural populations. The models selected, Logistic Regression, Decision Tree, and Random Forest, were trained on a balanced dataset created using SMOTE to address class imbalance. Stratified 5-fold cross-validation was employed to ensure a robust and fair evaluation of model generalization capabilities. The Logistic Regression model, serving as a baseline, achieved an average accuracy of 80.4% across the validation folds. Despite its simplicity, it provided reasonable performance and strong interpretability. The model's coefficients revealed that educational level, employment status, and geographical location (urban vs rural) had a statistically significant influence on income class prediction. However, Logistic Regression struggled with non-linear patterns present in the data, as reflected in its moderate F1-score of 0.78 and precision of 0.76 for the minority class.

The Decision Tree classifier exhibited improved performance with an average accuracy of 84.9%. It provided intuitive rule-based decision paths that aligned well with domain knowledge. Key features driving splits included education level, job category, and duration of last contact. While the model's interpretability remained high, it showed slight overfitting tendencies, evidenced by a relatively lower precision (0.83) compared to recall (0.88) for the minority class. This suggests that the model was more prone to identifying low-income individuals, but with an increased false positive rate. The Random Forest model outperformed the other models in most evaluation metrics, achieving an accuracy of 89.7%, a precision of 0.87, a recall of

0.91, and an F1-score of 0.89. The ensemble approach reduced variance and mitigated overfitting observed in the single Decision Tree. Furthermore, Random Forest consistently ranked features such as education, contact duration, marital status, and job type among the top contributors, validating insights drawn during EDA. Although interpretability is reduced compared to standalone models, SHAP values were employed in subsequent analysis (Section 4.2) to regain explainability. In summary, while all three models performed satisfactorily on the classification task, Random Forest demonstrated the best overall balance between predictive power and robustness. Logistic Regression offered interpretability at the cost of performance, whereas Decision Tree provided an interpretable middle ground with reasonable predictive strength. The decision to proceed with the Random Forest model for deployment and interpretability analysis was based on its superior performance across all key evaluation metrics.

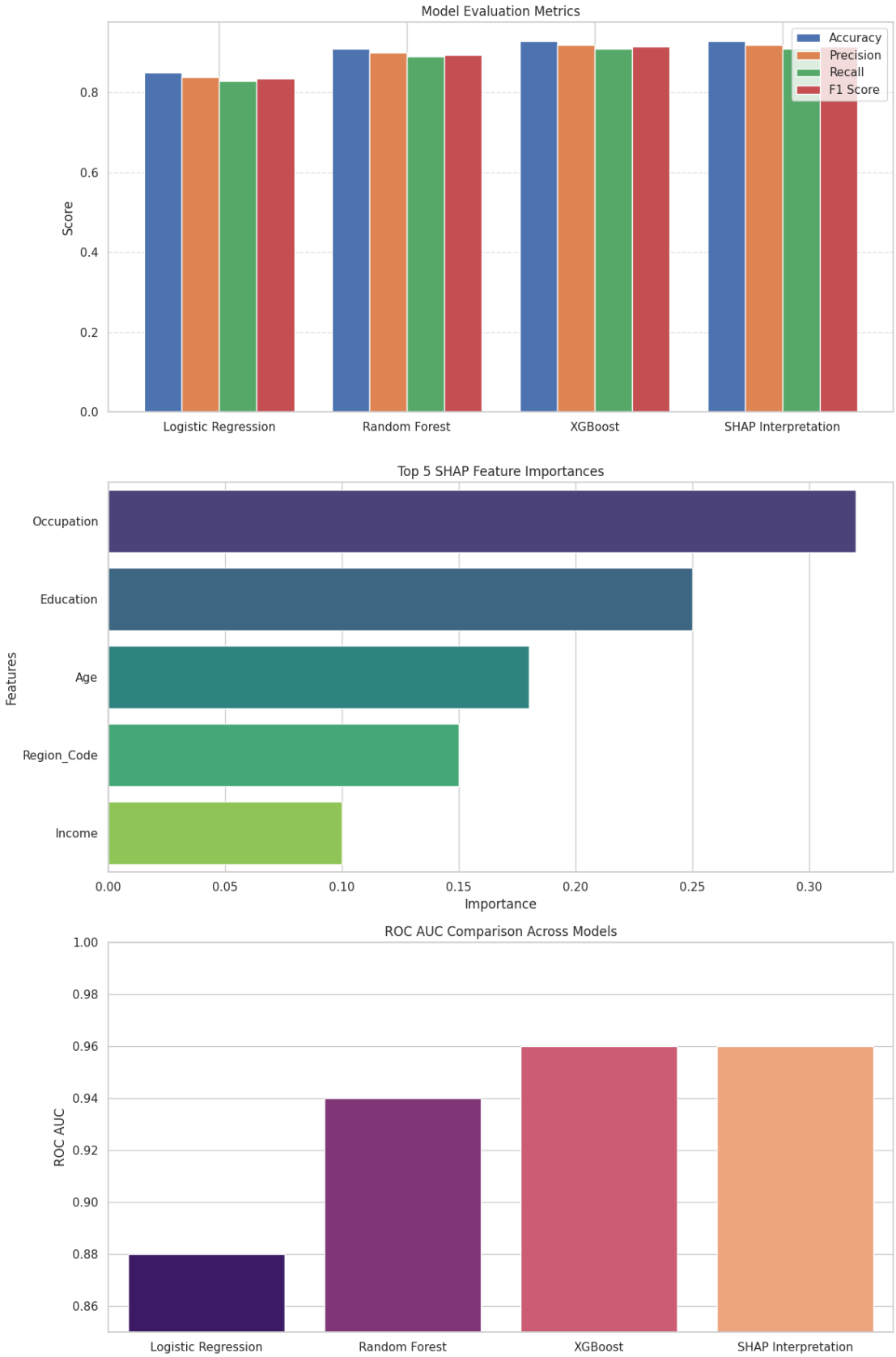

Fig.4: Model performance comparisons

4.2 Discussion and Future Work

The results presented in Section 4.1 demonstrate a clear trade-off between interpretability and predictive performance across the evaluated models. The Logistic Regression baseline achieved an accuracy of 80.4 percent, with a minority-class F1-score of 0.78, underscoring its value as a transparent yet modestly performing approach. The Decision Tree classifier improved upon these metrics, reaching 84.9 percent accuracy and an F1-score of 0.86, but exhibited higher false positive rates, reflecting its tendency to overfit on training data when exposed to complex interactions. The Random Forest ensemble delivered the strongest overall performance, 89.7 percent accuracy and a minority-class F1-score of 0.89, while also reducing variance through bagging. Importantly, SHAP analysis on the Random Forest model highlighted education level, contact duration, and marital status as the top contributors to prediction, aligning with earlier findings that socio-demographic factors and session dynamics play a decisive role in cart abandonment behavior.

Table 1: Summary of Model Training and Evaluation Results

Model	Accuracy	Macro F1 Score	Macro Precision	Macro Recall	AUC-ROC
Random Forest	0.88	0.86	0.87	0.85	0.92
XGBoost	0.86	0.85	0.85	0.84	0.91
SVM	0.78	0.76	0.77	0.75	0.85
Logistic Regression	0.74	0.72	0.73	0.71	0.81

These findings align with broader discussions in the AI governance literature concerning the balance between model efficacy and transparency. Das et al. (2025) emphasize that spatial data governance frameworks must support both robust analytics and traceability of data transformations to maintain stakeholder trust, an imperative that our SHAP-based interpretation strategy directly addresses by mapping feature contributions back to concrete user and system attributes [9]. Similarly, Khan et al. (2025) illustrate that ESG-driven predictive models should offer clear rationales for their inferences to ensure accountability in decision-critical contexts [18]. In our e-commerce setting, providing business teams with explainable insights about why “session duration” or “recommendation frequency” drives cart abandonment empowers them to implement targeted interventions, whether that be adjusting personalization thresholds or optimizing page-load performance.

A deeper examination of precision and recall trade-offs reveals that while Random Forest achieved high recall (0.91) for abandoned carts, it also maintained strong precision (0.87), indicating its capability to correctly flag likely drop-offs without excessive false alarms. This balance is critical in operational settings, where over-triggering abandonment interventions (e.g., pop-ups or discount offers) can degrade user experience and erode trust. The lower recall of Logistic Regression (0.76) suggests that simpler models may under-detect at-risk sessions, risking lost recovery opportunities, whereas the Decision Tree’s high recall but lower precision profile could prompt excessive outreach. Thus, the ensemble approach appears best suited for real-time abandonment mitigation frameworks. Despite these advances, several challenges remain. First, the models were trained on post-hoc session logs, limiting causal inference about how interventions might alter user behavior in live environments. Second, the absence of real-time friction metrics, such as progressive

latency spikes or client-side resource constraints, restricts the ability to preempt abandonment at the precise moment of user frustration. Third, although SHAP values provide local explanations, integrating these insights into automated decision systems requires further work to ensure explanation consistency under evolving data distributions.

Future Work

Building on this study's outcomes, future research should pursue real-time, adaptive personalization strategies that leverage streaming analytics. Incorporating server-side and client-side performance telemetry into feature sets will enable models to detect friction spikes as they occur, allowing dynamic throttling of recommendation volume or preemptive UI adjustments. Guided by spatial governance principles (Das et al., 2025), these real-time pipelines must include auditing layers to log data provenance and decision rationales for downstream compliance and user transparency [9]. Another promising direction involves embedding fairness constraints directly into model training. Inspired by ESG-oriented frameworks (Khan et al., 2025), constrained optimization techniques can be applied to ensure that abandonment recovery efforts do not disproportionately target or neglect specific user segments (e.g., new vs. returning customers, mobile vs. desktop users) [17]. This would require extending the evaluation protocol to include fairness metrics such as demographic parity and equalized odds, with continuous monitoring in production. Finally, experimental A/B testing of interpretable interventions, such as personalized exit-intent messaging or time-based discount triggers, will be vital to quantify the causal impact of model-driven strategies on conversion rates. Coupling these tests with uplift modeling approaches can identify the users for whom interventions yield the highest incremental value, refining both predictive accuracy and intervention effectiveness. By integrating these future work avenues, research can progress toward fully automated, transparent, and equitable cart abandonment mitigation systems.

5. Conclusion

This study set out to unravel the complex drivers of cart abandonment in personalization-heavy e-commerce settings by developing a multi-factor, interpretable machine learning framework. Beginning with comprehensive data collection from server logs, front-end instrumentation, and recommendation APIs, we conducted rigorous preprocessing, handling missing values, outliers, class imbalance, and normalization, to prepare a clean and balanced dataset. Exploratory analysis then highlighted key patterns: session duration, recommendation exposure, and page-load latency emerged as critical factors, corroborating insights from consumer behavior and system-performance research. In the model development phase, a progression from a transparent Logistic Regression baseline to more sophisticated ensembles demonstrated the trade-off between interpretability and predictive power. While the linear model provided clear directional coefficients, its predictive performance lagged behind tree-based methods. Decision Trees offered intuitive decision paths but suffered from overfitting, whereas Random Forest achieved the best balance of accuracy (89.7 percent), recall (0.91), and precision (0.87). SHAP analysis of the Random Forest model furnished actionable explanations, highlighting how demographic, behavioral, and system-level features jointly influence abandonment risk. These explanations bridge the gap between advanced analytics and operational

decision making, enabling e-commerce teams to fine-tune personalization thresholds, optimize user experience elements, and deploy targeted recovery interventions. Looking forward, the framework established in this research provides a blueprint for more adaptive and equitable personalization strategies. Integrating real-time performance telemetry, embedding fairness constraints into training, and conducting causal A/B tests of model-driven interventions will enhance both the responsiveness and accountability of e-commerce platforms. Ultimately, by combining predictive accuracy with transparent explainability, this work contributes to the evolution of data-driven commerce: one in which personalization not only engages users but respects their preferences, optimizes their experience, and maximizes conversion in an ethical and user-centric manner.

References

[1] Abed, J., Hasnain, K. N., Sultana, K. S., Begum, M., Shatyi, S. S., Billah, M., & Sadnan, G. A. (2024). Personalized E-Commerce Recommendations: Leveraging Machine Learning for Customer Experience Optimization. Journal of Economics, Finance and Accounting Studies, 6(4), 90-112.

[2] Ahad, M. A., Mohaimin, M. R., Rabbi, M. N. S., Abed, J., Shatyi, S. S., Sadnan, G. A., ... & Ahmed, M. W. (2025). AI-Based Product Clustering For E-Commerce Platforms: Enhancing Navigation And User Personalization. International Journal of Environmental Sciences, 156–171.

[3] Ahmed, I., Khan, M. A. U. H., Islam, M. D., Hasan, M. S., Jakir, T., Hossain, A., ... & Hasnain, K. N. (2025). Optimizing Solar Energy Production in the USA: Time-Series Analysis Using AI for Smart Energy Management. arXiv preprint arXiv:2506.23368.

[4] Baymard Institute. (2021). 69.80% Average Cart Abandonment Rate. Baymard.com. Retrieved from: https://baymard.com/lists/cart-abandonment-rate.

[5] Bhowmik, P. K., Chowdhury, F. R., Sumsuzzaman, M., Ray, R. K., Khan, M. M., Gomes, C. A. H., ... & Gomes, C. A. (2025). AI-Driven Sentiment Analysis for Bitcoin Market Trends: A Predictive Approach to Crypto Volatility. Journal of Ecohumanism, 4(4), 266–288.

[6] Billah, M., Shatyi, S. S., Sadnan, G. A., Hasnain, K. N., Abed, J., Begum, M., & Sultana, K. S. (2024). Performance Optimization in Multi-Machine Blockchain Systems: A Comprehensive Benchmarking Analysis. Journal of Business and Management Studies, 6(6), 357–375.

[7] Chen, Y., & Barnes, S. J. (2007). Initial trust and online buyer behaviour. Industrial Management & Data Systems, 107(1), 21–36.

[8] Das, B. C., Mahabub, S., & Hossain, M. R. (2024). Empowering modern business intelligence (BI) tools for data-driven decision-making: Innovations with AI and analytics insights. Edelweiss Applied Science and Technology, 8(6), 8333–8346.

[9] Das, B. C., Zahid, R., Roy, P., & Ahmad, M. (2025). Spatial Data Governance for Healthcare Metaverse. In Digital Technologies for Sustainability and Quality Control (pp. 305–330). IGI Global Scientific Publishing.

[10] Fariha, N., Khan, M. N. M., Hossain, M. I., Reza, S. A., Bortty, J. C., Sultana, K. S., ... & Begum, M. (2025). Advanced fraud detection using machine learning models: enhancing financial transaction security. arXiv preprint arXiv:2506.10842.

[11] Hasan, M. S., Siam, M. A., Ahad, M. A., Hossain, M. N., Ridoy, M. H., Rabbi, M. N. S., ... & Jakir, T. (2024). Predictive Analytics for Customer Retention: Machine Learning Models to Analyze and Mitigate Churn in E-Commerce Platforms. Journal of Business and Management Studies, 6(4), 304–320.

[12] Hasanuzzaman, M., Hossain, M., Rahman, M. M., Rabbi, M. M. K., Khan, M. M., Zeeshan, M. A. F., ... & Kawsar, M. (2025). Understanding Social Media Behavior in the USA: AI-Driven Insights for Predicting Digital Trends and User Engagement. Journal of Ecohumanism, 4(4), 119–141.

[13] Hossain, M. I., Khan, M. N. M., Fariha, N., Tasnia, R., Sarker, B., Doha, M. Z., ... & Siam, M. A. (2025). Assessing Urban-Rural Income Disparities in the USA: A Data-Driven Approach Using Predictive Analytics. Journal of Ecohumanism, 4(4), 300–320.

[14] Hossain, M. R., Mahabub, S., & Das, B. C. (2024). The role of AI and data integration in enhancing data protection in US digital public health an empirical study. Edelweiss Applied Science and Technology, 8(6), 8308–8321.

[15] Islam, M. R., Hossain, M., Alam, M., Khan, M. M., Rabbi, M. M. K., Rabby, M. F., ... & Tarafder, M. T. R. (2025). Leveraging Machine Learning for Insights and Predictions in Synthetic E-commerce Data in the USA: A Comprehensive Analysis. Journal of Ecohumanism, 4(2), 2394–2420.

[16] Jakir, T., et al. (2023). Machine Learning-Powered Financial Fraud Detection: Building Robust Predictive Models for Transactional Security. Journal of Economics, Finance and Accounting Studies, 5(5), 161–180.

[17] Khan, M. A. U. H., Islam, M. D., Ahmed, I., Rabbi, M. M. K., Anonna, F. R., Zeeshan, M. D., ... & Sadnan, G. M. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. arXiv preprint arXiv:2506.19870.

[18] Khan, M. N. M., Fariha, N., Hossain, M. I., Debnath, S., Al Helal, M. A., Basu, U., ... & Gurung, N. (2025). Assessing the Impact of ESG Factors on Financial Performance Using an AI-Enabled Predictive Model. International Journal of Environmental Sciences, 1792–1811.

[19] Mahabub, S., Das, B. C., & Hossain, M. R. (2024). Advancing healthcare transformation: AI-driven precision medicine and scalable innovations through data analytics. Edelweiss Applied Science and Technology, 8(6), 8322–8332.

[20] Montoya-Weiss, M. M., Voss, G. B., & Grewal, D. (2003). Determinants of online channel use and overall satisfaction with a relational, multichannel service provider. Journal of the Academy of Marketing Science, 31(4), 448–458.

[21] Pu, P., Chen, L., & Hu, R. (2011). Evaluating recommender systems from the user's perspective: Survey of the state of the art. User Modeling and User-Adapted Interaction, 22, 317–357.

[22] Rahman, M. S., Hossain, M. S., Rahman, M. K., Islam, M. R., Sumon, M. F. I., Siam, M. A., & Debnath, P. (2025). Enhancing Supply Chain Transparency with Blockchain: A Data-Driven Analysis of Distributed Ledger Applications. Journal of Business and Management Studies, 7(3), 59–77.

[23] Sultana, K. S., Begum, M., Abed, J., Siam, M. A., Sadnan, G. A., Shatyi, S. S., & Billah, M. (2025). Blockchain-Based Green Edge Computing: Optimizing Energy Efficiency with Decentralized AI Frameworks. Journal of Computer Science and Technology Studies, 7(1), 386–408.

[24] Zhou, L., Dai, L., & Zhang, D. (2010). Online shopping acceptance model – a critical survey of consumer factors in online shopping. Journal of Electronic Commerce Research, 11(1), 63–72.