Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce

MD Tanvir Islam[1], Eftekhar Hossain Ayon[2]✉ Bishnu Padh Ghosh[3], MD, Salim Chowdhury[4], Rumana Shahid[5], Aisharyja Roy puja[6], Sanjida Rahman[7], Mohammad Shafiquzzaman Bhuiyan[8] and Tuan Ngoc Nguyen[9]

[1]*Department of Computer Science, Monroe College, New Rochelle, New York, US*
[2]*Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA*
[3]*School of Business, International American University, Los Angeles, California, USA*
[4]*College of Graduate and Professional Studies Trine University, USA*
[5]*Department of Management of Science and Quantitative Methods, Gannon University, USA*
[6]*Department of Management Science and Quantitative Methods, Gannon University, USA*
[7]*Department of Public Administration, Gannon University, Erie, PA, USA*
[8]*Department of Business Administration, Westcliff University, Irvine, California, USA*
[9]*VNDirect Securities, 97 Lo Duc, Hai Ba Trung, Hanoi, Vietnam*

Corresponding Author: Eftekhar Hossain Ayon, **E-mail**: ayon001@gannon.edu

| ABSTRACT

A thorough comparison of several machine learning methods is provided in this paper, including gradient boosting, AdaBoost, Random Forest (RF), XGBoost, Artificial Neural Network (ANN), and a unique hybrid framework (RF-XGBoost-LR). The assessment investigates their efficacy in real-time sales data analysis using key performance metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2 score. The study introduces the hybrid model RF-XGBoost-LR, leveraging both bagging and boosting methodologies to address the limitations of individual models. Notably, Random Forest and XGBoost are scrutinized for their strengths and weaknesses, with the hybrid model strategically combining their merits. Results demonstrate the superior performance of the proposed hybrid model in terms of accuracy and robustness, showcasing potential applications in supply chain studies and demand forecasting. The findings highlight the significance of industry-specific customization and emphasize the potential for improved decision-making, marketing strategies, inventory management, and customer satisfaction through precise demand forecasting.

| KEYWORDS

Revolutionizing Retail; Hybrid Machine Learning Approach; Global Commerce

| ARTICLE INFORMATION

ACCEPTED: 15 December 2023 **PUBLISHED:** 02 January 2024 **DOI:** 10.32996/jcsts.2024.6.1.4

1. Introduction

In the ever-evolving landscape of global commerce, major retailers play a pivotal role in shaping market dynamics and consumer expectations. This study delves into the operations of a prominent corporate entity functioning as a major retailer renowned for its robust global presence and diverse product portfolio. Central to its strategy is a distinctive pricing approach that strategically reduces grocery prices in markets where it operates. Beyond the retail realm, the company's market power extends to include ancillary services, further solidifying its standing as a dynamic force in the global market. At the heart of this investigation lies a comprehensive dataset derived from the operations of this retail giant. This dataset serves as the cornerstone for the development and evaluation of a cutting-edge hybrid model, aiming for real-time analysis of sales data and a substantial enhancement in forecasting accuracy. By leveraging the wealth of information encapsulated in this dataset, the study explores the intricate relationship between the company's market strategies, product diversification, and its dynamic position in the global marketplace.

Machine learning (ML) techniques have gained widespread adoption across diverse domains, offering solutions to challenges posed by large and intricate datasets. These techniques use complex algorithms to find relevant patterns in large and varied datasets—a task that would be nearly impossible for a human analyst to accomplish. The main goal of machine learning is inductive inference, which is the process of creating general models from particular empirical data. The constant creation of new algorithms and the increase in data availability at lower computational costs have driven advancements in this field.

When combined with predictive analysis, machine learning improves customer interaction and allows for more accurate demand projections than with conventional techniques. Retail chains perform better overall when ML techniques are used to handle complex correlations between multiple causal elements, especially nonlinear relational demand patterns. Although popular predictive models for demand forecasting include auto-regressive integrated moving average (ARIMA) and auto-regressive integrated moving average with exogenous variables (ARIMAX), newly developed machine learning (ML) algorithms like artificial neural networks (ANN), support vector machines (SVM), and regression trees have outperformed more conventional techniques. This study's main goals are to (1) investigate the machine learning models used in forecasting and (2) compare certain machine learning and hybrid models for sales forecasting in a US-based retail company.

This study conducts a comprehensive comparison of various ML models for retail demand forecasting, including random forest (RF), artificial neural network (ANN), gradient boosting (GB), adaptive boosting (AdaBoost), and extreme gradient boosting (XGBoost). The performance of these models is benchmarked against a proposed hybrid model that combines RF, XGBoost, and linear regression (LR). Diverse performance metrics such as mean squared error (MSE), R2 score, and mean absolute error (MAE) are considered to ensure an accurate evaluation. The methodologies' advantages and limitations are explored, along with potential avenues for future performance enhancements. The historical sales data of a prominent US-based multinational retail company form the basis for the forecasting analysis, considering various stores across the USA that cater to a wide range of consumer needs.

The intricate interplay of market dynamics, pricing strategies, and global competitiveness underscores the significance of understanding and optimizing forecasting methodologies for such retail giants. As consumer preferences shift and market forces continue to evolve, the ability to accurately predict and respond to demand becomes a critical factor in ensuring sustainable growth and strategic decision-making. This study, therefore, seeks to unravel the complexities of this major retailer's operations, shedding light on how its distinct pricing strategy and diversified product offerings contribute to its success. Through the lens of a sophisticated hybrid model, the research aims to not only analyze historical sales data but also to provide valuable insights that could shape the future trajectory of this retail giant in the ever-changing landscape of global commerce.

2. Literature Review

In their recent study, Namburu et al. (2022) underscore the critical importance of effective pricing strategies, particularly in the dynamic landscape intensified by the surge in online shopping during the COVID-19 pandemic. Conducting an extensive literature review, the paper delves into existing research on pricing strategies and market analysis, emphasizing the role of machine learning models in personalized dynamic pricing and market trend prediction. The study aims to provide scalable pricing solutions, aid retailers in swift decision-making, and offer repeatable pricing strategies. The methodology involves meticulous data preprocessing, feature engineering, exploratory data analysis, and the creation of novel features to enhance the performance of machine learning models. The introduction of ensemble learning techniques, such as XGBoost, LightGBM, and CatBoost, sets the stage for the innovative X-NGBoost algorithm—a hybrid approach integrating XGBoost with natural gradient boosting, demonstrating superior speed and accuracy. The paper thoroughly compares existing algorithms based on various criteria, concluding that X-NGBoost outperforms its counterparts, which is evident in the lower root-mean-square error (RMSE). The significance of the proposed methodology for small-scale retailers is highlighted in the conclusion, suggesting future work involving the extension of ensemble techniques to predict pricing solutions across multiple e-commerce platforms. Gupta et al. (2014) embarked on a comprehensive project to develop a framework and effective methodologies, utilizing robust machine learning algorithms to optimize customer decision-making regarding well-priced purchases on e-commerce platforms. Focusing on inventory-centric e-commerce companies, the study demonstrates adaptability for application in online marketplaces without physical inventories. Leveraging statistical and machine learning models, the research aims to forecast purchase decisions through adaptive or dynamic pricing strategies. Various data sources, encompassing visit attributes, visitor characteristics, purchase history, web data, and context understanding, form the foundation of this framework. In contrast to individual buyers, the study emphasizes customer segments to enhance the accuracy of purchase predictions, integrating web mining, big data technologies, and machine learning algorithms into a cohesive solution landscape to address the complexities of optimizing pricing decisions in e-commerce.

3. Methodology

The efficiency of several machine learning techniques, such as XGBoost, Random Forest (RF), Artificial Neural Network (ANN), gradient boosting, AdaBoost, and a unique hybrid framework (RF-XGBoost-LR), was thoroughly compared in the study. Important metrics like mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R2) score were used in performance evaluations. Notably, ANN is an example of a deep-learning strategy, whereas XGBoost, RF, gradient boosting, and AdaBoost are ensemble techniques based on decision trees. The study emphasized the three nodes that make up a decision tree (DT): the root, internal, and leaf nodes. Decision tree algorithms, guided by simple principles, traverse from the root node through internal nodes, culminating in the leaves. The implementation was carried out using Python, with data handling facilitated by Pandas and NumPy versions. Model training utilized XGBoost and Scikit-learn.

3.1 Proposed Methodology

The study introduces a hybrid model, RF-XGBoost-LR, and compares its performance with that of individual models. A fundamental distinction exists between bagging and boosting methodologies. Bagging aims to reduce prediction variance by creating additional training data through repeated random sampling from the dataset, generating multiple subsets. On the other hand, boosting adjusts the weight of observations based on their classification accuracy in previous iterations. In contrast to bagging's uniform sample selection for training datasets, boosting employs a non-uniform sampling approach. Samples with higher weights, often indicating misclassifications, have a greater likelihood of being chosen. Consequently, boosting algorithms tend to focus on samples that previous models have struggled to classify accurately.

Random Forest (RF) functions as an ensemble technique that amalgamates the results from a multitude of regression trees to generate a cohesive prediction. At the heart of RF lies the concept of bagging, a method where a random sample is drawn from the training data to construct an individual regression tree. This subset, referred to as a bootstrap sample, is selected with replacement, introducing the possibility of reselecting any data point that had been previously chosen. The process of creating a bootstrap sample involves the random selection of N data points from the dataset, and subsequently, these selected points are replaced in the dataset. This methodology allows for variability in the training process and enhances the robustness of the overall Random Forest model by leveraging diverse subsets of data in the construction of individual trees, ultimately contributing to a more reliable and generalized predictive performance.

XGBoost, short for 'extreme gradient boosting,' represents a potential advancement in gradient boosting techniques, demonstrating notable improvements. Renowned for its ability to enhance performance and address real-world challenges at scale with minimal resource utilization, XGBoost operates as a parallel tree model, extending the principles of the gradient boosting model. Its framework employs the tree ensemble method, consisting of a sequence of Classification and Regression Trees (CART). While XGBoost encompasses distinctive features, certain aspects, such as the second-order Taylor expansion and integrated normalization algorithms, share similarities with gradient-boosted decision Trees (GBDT). Notably, XGBoost models exhibit the advantage of efficient scalability across diverse scenarios, outperforming existing prediction models while demanding fewer resources. The incorporation of parallel and distributed computation within XGBoost accelerates the model learning process, facilitating rapid exploration of model variations.

3.2 Hybrid Model

The hybrid model, denoted as RF-XGBoost-LR, leverages the strengths of Random Forest (RF), which generates parallel decision trees to mitigate overfitting issues, leading to enhanced accuracy through variance reduction. In RF, distinct decision trees are created for each copy of the original training data, contributing to improved generalization. Despite the widespread popularity of random forest, it grapples with both conceptual and practical limitations. The adaptive learning mechanism of random forest exhibits inherent weaknesses in minimizing training errors, as each tree is learned independently, and the synergistic insights from other trees are not fully harnessed during training. Consequently, this independence among trees can result in diminished model performance. On the other hand, XGBoost employs a sequential approach to combine multiple weak learners iteratively, leading to continuous improvement in overall model performance.

XGBoost, functioning as a boosting technique, capitalizes on parallel processing by executing the model across multiple CPU cores. However, it grapples with the well-known issue of overfitting, a common concern in boosting methodologies, including MARTs, which stand for multiple additive regression trees. This problem intensifies when there are few trees available at the beginning of the iteration, causing all the trees to have a major influence on the model. When a model is overfitted with training data, it becomes less generalizable and performs unreliablely when used with fresh measurements. High variance and low bias estimators are common signs of overfitting, which introduces complexity and may improve the model's performance on training data but impairs accurate point predictions in subsequent scenarios. When new data from the underlying population is applied to prediction results, overfitting results in an overly optimistic representation of the predictions.

Hybrid models refer to amalgamations of two or more individual machine learning or soft computing models, aiming to attain increased versatility and capability compared to a singular model. These hybrid models typically incorporate both prediction and optimization aspects to enhance accuracy. The development of hybrid models is primarily driven by two key objectives: firstly, to mitigate the risk associated with inaccurate predictions that a single forecast might yield under certain conditions, and secondly, to enhance the overall performance beyond that achievable by individual models.

The development of the hybrid model is intricately designed to capitalize on the inherent strengths and simultaneously address the weaknesses present in the individual models under consideration. Within the scope of this study, an innovative hybrid machine learning (ML) model is introduced, integrating both the random forest regressor—a proficient bagging technique—and the XGBoost regressor—an effective boosting technique.

The primary objective behind formulating the hybrid model is to tackle the inherent limitations observed in both the Random Forest (RF) and XGBoost models. Specifically, the random forest component of the hybrid model is strategically employed to combat the overfitting challenge associated with XGBoost. By adeptly reducing model variance without introducing a concurrent increase in model bias, the random forest model presents a nuanced solution. This approach implies that while the overfitting issue may manifest in the prediction of an individual regression tree, it can be effectively mitigated when considering the aggregate prediction derived from multiple regression trees.

Despite the notable advantages, the random forest model may exhibit certain shortcomings in terms of reducing training error stemming from the autonomous training of multiple regression trees. However, the hybrid model strategically integrates XGBoost to overcome this limitation. XGBoost, through its sequential training of decision trees, provides a complementary mechanism to enhance the overall performance of the hybrid model. This sequential training methodology ensures that the hybrid model not only leverages the collective strengths of both random forest and XGBoost but also navigates the intricacies associated with independent training, offering a comprehensive and robust solution to the challenges posed by individual models.

4. Result and Discussion

The corporate organization is a consumer products retailer that operates globally, concentrating on supervising a variety of locations outside of the United States, such as supercenters, supermarkets, hypermarkets, warehouse clubs, and cash and carries. With its headquarters located in Bentonville, Arkansas, the company was founded in 1945 and has grown to become one of the biggest retailers globally, achieving year-over-year revenue growth. With operations in more than 25 countries, it is a global powerhouse that operates grocery stores, supermarkets, hypermarkets, department stores, and discount stores with a unique focus on providing goods at the lowest possible cost.

The broad range of products offered by this retail behemoth includes everything from jewelry and baby supplies to office supplies, electronics, books, movies, and music, as well as furniture and even pharmaceuticals. Statistics show that the company can significantly lower grocery prices; in markets where it establishes a presence, prices typically decrease by 5–10%. With this pricing strategy, the business not only establishes itself as a strong rival but also emphasizes its dedication to offering customers affordable solutions.

Additionally, by consistently providing products at the lowest prices, the company maintains a competitive edge over suppliers and rivals by leveraging its strong market power. Due to the company's diversification into ancillary services like fuel, gift cards, banking services, money orders, prepaid cards, and wire transfers, it has a strategic advantage that goes beyond the retail sector. This retail behemoth continues to be a dynamic force in the global market, influencing industry trends and setting consumer expectations through constant innovation and adaptation of its business model.

This study has made use of data from a retailer that is well-known for satisfying a wide range of consumer needs with a wide selection of products. The dataset includes sales information for 45 stores and 99 departments from various regions in the United States over a three-year period. The dataset comprises multiple attributes that offer specifics about individual stores, including number, size, department, week date, and average temperature. Additionally, region-specific data is provided, such as the Consumer Price Index (CPI), holiday weeks and average fuel price.

In machine learning, normalization is an essential preprocessing step that is critical to improving the effectiveness of the learning process. The goal is to shorten the learning curve by normalizing the data, particularly when working with large datasets. Min-Max normalization is a technique that uses a linear transformation to change the original dataset into a specified interval. This normalization method's primary benefit is its capacity to maintain all correlations among data points, guaranteeing that significant insights are retained throughout the learning process.

In the evaluation of forecasting models, various metrics are employed, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and the R2 value, as elaborated in the subsequent sections. Mean Absolute Error serves as an average metric to quantify the errors within a set of predictions. Being an absolute measure, it disregards the direction of errors, treating both positive and negative errors with equal weight. The computation of MAE is straightforward, involving the summation of the absolute values of errors to obtain the 'total error,' which is then divided by the total number of observations. This metric provides a comprehensive assessment of the accuracy of predictions, focusing on the magnitude of errors while maintaining a balanced consideration of individual errors across the entire dataset. The simplicity and transparency of the MAE calculation make it a valuable tool in evaluating the predictive performance of forecasting models, shedding light on the overall precision of the predictions without being influenced by the directionality of errors.

The coefficient of determination, or R2 Score, is a metric that expresses how much of the variance in a dependent variable can be explained by a particular model. When evaluating the dispersion of data points surrounding a fitted regression line, this metric is crucial. When applied to comparable datasets, high R2 values indicate little differences between the expected and actual data. On a scale from 0 to 1, the R2 score measures the correlation between the predicted and actual data. For example, an R2 value of 0.8 means that the variation in the independent variable under investigation can account for 80% of the variability in the dependent variable. This metric gives useful information about a model's effectiveness by providing a percentage-based understanding of how well it captures the variability in the observed data.

R2 is primarily used to evaluate the goodness of fit that a given model provides for the observed values. Furthermore, it is imperative to comprehend errors, and this was accomplished by employing metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE). By giving large errors resulting from value squaring more significance, the Mean Squared Error serves the purpose of prioritizing outliers. However, when evaluating forecasts in the same unit as the original series, Mean Absolute Error is used to provide information about the average expected error from the forecast. The study compares different models using the performance parameters (MSE, MAE, R2) that were previously mentioned. The findings are displayed in Table 1. This method makes it easier to evaluate the models thoroughly based on how well they fit the observed values and the size of the errors.

Model	**MSE**	**MAE**	R^2
Random forest	5.82e-05	0.0027	0.9251
XGBoost	4.72e-05	0.0023	0.9447
Gradient boosting	8.42e-05	0.0033	0.9016
AdaBoost	6.48e-05	0.0027	0.9208
Artificial neural network	5.46e-04	0.0139	0.3858
RF-XGBoost-LR (hybrid)	**4.89e-04**	**0.0025**	**0.9651**

Various performance metrics were used to evaluate and contrast each model's efficacy. With an R2 score of 0.9651, an MAE of 0.0025, and an MSE of 4.8932e-04, the recommended forecasting approach outperformed the other benchmark methods overall when three metrics were considered.

The maximum depth of 28 and 175 estimators were used to set the parameters for the RF model. A maximum depth of 25 and 125 estimators were selected for the gradient boosting model. The base estimator in the AdaBoost model was a decision tree with a depth of 25. The artificial neural network (ANN) had five layers in its architecture, with 10, 12, 24, 12, and 10 neurons in each layer. Each layer's activation function was set to'relu,' and the 'adam' optimizer was applied. Training took place in 500 epochs with a 256-batch size. 150 estimators with a maximum depth of 25 were chosen for the XGBoost model. RF and XGBoost were combined in the RF-XGBoost-LR model, with the predictions from these models serving as input to a logistic regression (LR) model.

Finding the optimal configuration parameters is a challenge in the hyperparameter optimization stage of machine learning models. As such, using random values that fall into the appropriate range of algorithm parameters can result in better optimization results.

The RF and XGBoost models' outputs were fed into an LR model, which was selected for its ease of use in producing final predictions. To reduce the possibility of overfitting in the hybrid model, a more sophisticated model—such as gradient boosting—was not chosen for the final layer. By addressing the drawbacks of the RF and XGBoost models, the hybrid model helped to create a cohesive and reliable solution.

The proposed hybrid model has the potential to enhance studies related to supply chain dynamics and can contribute to the extension of research in demand forecasting. The robust performance exhibited by this framework broadens its applicability, offering benefits to retailers, wholesalers, and various industries. However, customization tailored to specific domains is crucial, necessitating adequate domain knowledge for diverse applications across industries. Different sectors may possess unique product properties that, when integrated into the forecasting framework, can significantly enhance its performance, thereby warranting exploration as a potential research avenue.

The study's conclusions show that, in comparison to standalone machine learning models, the recommended hybrid model greatly improves forecasting accuracy. Through the combined efforts of random forest and XGBoost, the model successfully addresses problems such as overfitting and training errors in linear regression analysis, resulting in forecast values that are closely aligned with actual sales figures. Therefore, better forecasting can help industry decision-makers with marketing strategies, inventory turnover, capacity planning, supply chain cost reduction, and customer satisfaction. Overall, forecasting accuracy can be increased. Through the reduction of the bullwhip effect and the encouragement of efficient inventory management, the accuracy of the demand forecasting approach can improve the performance of the supply chain.

5. Conclusion

In order to analyze sales data in real-time, this study presents a novel hybrid machine learning (ML) model that combines XGBoost, Random Forest (RF), and Logistic Regression (LR). A retail company's sales data with a variety of attributes is trained to present a more sophisticated model with higher accuracy. A hybrid strategy is suggested to address the shortcomings in the RF and XGBoost models. The dataset is first normalized, and then it is separately trained and tested in the RF and XGBoost models. After that, the predictions made by each of these models are combined to create a new dataset, which is used as input by the LR model to produce its final predictions.

By reducing variance and strengthening robustness to outliers, the combination of XGBoost and RF models is found to improve dataset accuracy. This leads to better predictive ability and less vulnerability to overfitting. Three metrics are used in the study: R2 scores, Mean Squared Error (MSE), and Mean Absolute Error (MAE). The results show that RF-XGBoost-LR, the proposed hybrid model, performs better than RF, ANN, gradient boosting, AdaBoost, and XGBoost (MAE = 0.0025, MSE = 4.8932e-05, and R2 score = 0.9551). With an R-squared of 96.51%, the model appears to account for a sizable amount of the included data and variables.

The corporate entity, functioning as a major retailer, showcased a robust global presence and a diversified product portfolio, emphasizing a distinctive pricing strategy that reduced grocery prices in markets where it operated. Leveraging its market power, the company extended its offerings to include ancillary services, solidifying its position as a dynamic force in the global market. The dataset from this retail giant served as the foundation for the proposed hybrid model, contributing to its real-time analysis of sales data and enhancing forecasting accuracy.

The hybrid model, RF-XGBoost-LR, not only overcame the limitations inherent in individual models but also demonstrated superior performance across multiple performance metrics. The study's findings have broad implications for supply chain-related studies, offering a robust forecasting tool that can be applied across industries. The precision of the hybrid model aids decision-makers in formulating better marketing strategies, optimizing inventory turnover, planning capacity effectively, reducing supply chain costs, and enhancing overall customer satisfaction.

While the study provides valuable insights, it is essential to acknowledge the need for domain-specific customization for diverse applications. The hybrid model's adaptability and accuracy underscore its potential for widespread implementation, but a nuanced understanding of industry-specific nuances remains crucial. Future research endeavors could explore the integration of unique product properties from different sectors to further enhance the hybrid model's performance. In essence, the proposed hybrid model represents a promising advancement in the field of demand forecasting, with the potential to revolutionize supply chain dynamics and decision-making processes across various industries.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References:

[1] Ahmed, A. H., Ahmad, S., Sayed, M. A., Ayon, E. H., Mia, T., & Koli, T. (2023). Predicting the Possibility of Student Admission into Graduate Admission by Regression Model: A Statistical Analysis. *Journal of Mathematics and Statistics Studies*, *4*(4), 97-105.

[2] Abu S., Tayaba, M., Islam, M. T., Eyasin U P., Tuhin M., Eftekhar H A., Nur N & Bishnu P G. (2023). Parkinson's Disease Detection through Vocal Biomarkers and Advanced Machine Learning Algorithms. *Journal of Computer Science and Technology Studies*, *5*(4), 142–149. https://doi.org/10.32996/jcsts.2023.5.4.14

[3] Gupta, R., & Pathak, C. (2014). A machine learning framework for predicting purchase by online customers based on dynamic pricing. *Procedia Computer Science*, *36*, 599-605.

[4] Haque, M. S., Amin, M. S., Ahmad, S., Sayed, M. A., Raihan A and Hossain, M. A. (2023). Predicting Kidney Failure using an Ensemble Machine Learning Model: A Comparative Study, 2023 10th International Conference on Electrical Engineering, Computer Science, and Informatics (EECSI), Palembang, Indonesia. 31-37, doi: 10.1109/EECSI59885.2023.10295641.

[5] Ho, W. K., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, *38*(1), 48-70.

[6] Khan, R. H., Miah, J., Arafat, S. M., Syeed, M. M., & Ca, D. M. (2023). Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks. *arXiv preprint arXiv:2310.17729*.

[7] Khan R. H and Miah, J. (2022). Performance Evaluation of a new one-time password (OTP) scheme using stochastic petri net (SPN), IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA. 407-412, doi: 10.1109/AIIoT54504.2022.9817203.

[8] Khan, R. H., Miah, J., Rahman M and Tayaba, M. (2023). A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA. 647-652, doi: 10.1109/CCWC57344.2023.10099106.

[9] Khan, R. H., Miah, J., Rahat, M. A. R., Ahmed, A. H., Shahriyar M. A and Lipu, E. R. (2023). A Comparative Analysis of Machine Learning Approaches for Chronic Kidney Disease Detection, 2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Malang City, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICEEIE59078.2023.10334765.

[10] Khan, R. H., Miah, J., Nipun, S. A. A., Islam, M., Amin M. S and Taluckder, M. S. (2023). Enhancing Lung Cancer Diagnosis with Machine Learning Methods and Systematic Review Synthesis, 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Malang City, Indonesia. 1-5, doi: 10.1109/ICEEIE59078.2023.10334739.

[11] Li, X., Xu, M., & Yang, D. (2021). Cause-related product pricing with regular product reference. *Asia-Pacific Journal of Operational Research*, *38*(04), 2150002.

[12] Modak, C., Shahriyar, M. A., Taluckder, M. S., Haque M. S and Sayed, M. A. (2023). A Study of Lung Cancer Prediction Using Machine Learning Algorithms, 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), Yogyakarta, Indonesia, 2023. 213-217, doi: 10.1109/ICE3IS59323.2023.10335237.

[13] Miah, J., & Khan, R. H. (2019, November). Service Development of Smart Home Automation System: A Formal Method Approach. In *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* (161-167).

[14] Miah, J., Ca, D. M., Sayed, M. A., Lipu, E. R., Mahmud, F., & Arafat, S. M. Y. (2023). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction. arXiv preprint, 2311.00517.

[15] Mridha, M. M., Hossain, M. S., Chowdhury, M. S., Al-Imran, M. & Debnath, D. (2023). Factors influencing consumers' attitude towards online shopping in Rangpur City. *International Multidisciplinary Research Journal* 16-26. [Available from: https://doi.org/10.25081/imrj.2023.v13.8324]

[16] Miah, J., Khan, R. H., Ahmed S and Mahmud, M. I. (2023). A comparative study of Detecting Covid 19 by Using Chest X-ray Images– A Deep Learning Approach, 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, pp. 0311-0316, doi: 10.1109/AIIoT58121.2023.10174382.

[17] Miah, J., Cao, D. M., Sayed, M. A., & Haque, M. S. (2023). Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks. *arXiv preprint arXiv:2310.18237*.

[18] Miah, J., Haque, M. S., Cao, D. M., & Sayed, M. A. (2023). Enhancing Traffic Density Detection and Synthesis through Topological Attributes and Generative Methods. *Journal of Computer Science and Technology Studies*, *5*(4), 69–77. https://doi.org/10.32996/jcsts.2023.5.4.8

[19] Namburu A, Selvaraj Varsha M. (2022). Product pricing solutions using hybrid machine learning algorithm. Innov Syst Softw Eng. 2022 Jul 25:1-12. Doi: 10.1007/s11334-022-00465-3. Epub ahead of print. PMID: 35910813; PMCID: PMC9309595.

[20] Peiseler, F., Rasch, A., & Shekhar, S. (2022). Imperfect information, algorithmic price discrimination, and collusion. *The Scandinavian Journal of Economics*, *124*(2), 516-549.

[21] Syeed, M. M., Khan, R. H., & Miah, J. (n.d). Agile Fitness of Software Companies in Bangladesh: An Empirical Investigation.

[22] Sarkar, M., Ayon, E. H., Mia, M. T., Ray, R. K., Chowdhury, M. S., Ghosh, B. P., Al-Imran, M., Islam, M. T., Tayaba, M., & Puja, A. R. (2023). Optimizing E-Commerce Profits: A Comprehensive Machine Learning Framework for Dynamic Pricing and Predicting Online Purchases. *Journal of Computer Science and Technology Studies*, *5*(4), 186–193. https://doi.org/10.32996/jcsts.2023.5.4.19