comien

International Journal on Computational Engineering

Journal Homepage www.comien.org/index.php/comien

Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications

Md Murshid Reja Sweet [a], Md Arif [a], Aftab Uddin [b], Kazi Shaharair Sharif [c], Mazharul Islam Tusher [d], Suniti Devi [a], Md Parvez Ahmed [e], Maniruzzaman Bhuiyan [f], Md Habibur Rahman [g,*], Abdullah Al Mamun [h], Tauhedur Rahman [i], Md Asaduzzaman [j], Md Jamil Ahmmed [k], Md Ariful Islam Sarkar [l]

[a] *Department of Management Science and Quantitative Methods, Gannon University, USA*
[b] *Fox School of Business & Management, Temple University, USA*
[c] *Department of Computer Science, Oklahoma State University, USA*
[d] *Department of Computer Science, Monroe College, New Rochelle, New York, USA*
[e] *Master of Science in Information Technology, Washington University of Science and Technology, USA*
[f] *Satish & Yasmin Gupta College of Business, University of Dallas, Texas, USA*
[g] *Department of Business Administration, International American University, Los Angeles, California, USA*
[h] *Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA*
[i] *Dahlkemper School of Business, Gannon University, USA*
[j] *Department of Business Administration, Northern University, Bangladesh*
[k] *Department of Project Management and Business Analytics, Francis College, USA*
[l] *Department of Business Administration, Stamford University Bangladesh*

Corresponding author: [*]rahman010@gannon.edu

***Abstract*— This study evaluates the effectiveness of statistical and machine learning models in credit risk assessment, comparing traditional methods like logistic regression with advanced techniques such as Decision Trees, SVM, Neural Networks, and GBM. The results demonstrate that Neural Networks and GBM achieve the highest predictive accuracy (0.88 and 0.87, respectively), excelling in capturing complex borrower behaviors. Conversely, logistic regression, though more interpretable, shows a lower accuracy of 0.75, highlighting its limitations. The paper underscores the balance needed between model complexity and interpretability, especially in regulatory settings, and provides practical insights for optimizing credit risk assessment models.**

***Keywords*— Borrower behavior; model interpretability; credit; neural networks; Gradient Boosting Machines (GBM)**

Manuscript received 15 Feb. 2024; revised 29 Apr. 2024; accepted 12 Aug. 2024. Date of publication 30 Sep. 2024.

I. Introduction

Credit risk assessment is a fundamental process for financial institutions, aimed at evaluating the likelihood of a borrower defaulting on their financial obligations. This assessment plays a critical role in decision-making processes, including loan approvals, pricing, and portfolio management. The accurate prediction of credit risk is not only essential for the profitability of financial institutions but also for maintaining financial stability and compliance with regulatory requirements.

Traditionally, credit risk assessment has relied heavily on statistical methods, such as logistic regression and linear discriminant analysis. These models are prized for their simplicity and interpretability, providing straightforward probabilistic frameworks to estimate the likelihood of default based on borrower characteristics. However, the complexity of financial markets and borrower behavior has revealed limitations in these methods, particularly their struggle with non-linear relationships and interactions within the data.

The emergence of machine learning techniques has revolutionized the field by introducing advanced algorithms capable of uncovering complex patterns from large datasets. Decision Trees, Support Vector Machines (SVM), and Neural Networks [1], [2] represent some of the leading methodologies, each bringing unique strengths to the task of

credit risk assessment. Decision Trees, for example, offer intuitive visualizations of decision-making processes but are prone to overfitting. Ensemble methods like Random Forest and Gradient Boosting Machines (GBM) [3], [4] mitigate this issue by averaging predictions from multiple trees, thereby enhancing model robustness and accuracy. SVMs, known for their efficiency in high-dimensional spaces, and Neural Networks, celebrated for their ability to model non-linear relationships, further extend the toolkit available for assessing credit risk.

This paper explores the integration of these statistical and machine learning methodologies in credit risk assessment. We provide a comprehensive review of both traditional and advanced models, emphasizing their methodologies, strengths, and weaknesses. Furthermore, we evaluate these models using a real-world dataset, analyzing their predictive performance based on various metrics. The study concludes with a discussion on practical applications, challenges, and future directions in credit risk modeling, aiming to provide insights into the effective utilization of these tools in the financial sector.

A. Traditional Statistical Methods

Traditional statistical methods have long been the cornerstone of credit risk assessment. Logistic regression, one of the most widely used models, provides a probabilistic framework for predicting binary outcomes, such as default or non-default. The model estimates the likelihood of default based on borrower characteristics, such as income, employment status, and credit history. Other methods, like linear discriminant analysis, have also been employed to distinguish between default and non-default cases. These models are favored for their interpretability and simplicity but may struggle with non-linearity and complex interactions in the data.

B. Evolution to Machine Learning

The advent of machine learning has revolutionized credit risk assessment by introducing algorithms capable of capturing complex patterns in large datasets. Decision trees, for instance, offer a visual and interpretable way of segmenting data based on feature importance. However, they are prone to overfitting. To mitigate this, ensemble methods like Random Forest and Gradient Boosting Machines (GBM) combine multiple decision trees to enhance predictive accuracy and generalization.

Support Vector Machines (SVM) [5], [6] provide another powerful tool, especially in handling high-dimensional data. SVMs use hyperplanes to separate classes, aiming to maximize the margin between them. Despite their robustness, SVMs can be computationally intensive and less interpretable.

Neural networks, particularly deep learning architectures, have gained prominence for their ability to model non-linear relationships. With multiple hidden layers, these networks can learn intricate patterns from the data. However, they require substantial computational resources and large amounts of data for training.

C. Comparative Studies and Practical Applications

Several studies have compared the performance of these models in credit risk assessment. Mohammad et al [7] conducted a comprehensive benchmarking study, concluding that ensemble methods and neural networks generally outperform traditional statistical models in terms of predictive accuracy. However, the choice of model often depends on the specific context, including data availability, regulatory requirements, and the need for model interpretability.

Practical applications of these models span various financial institutions, from banks to credit card companies. They are used not only for loan approval decisions but also for pricing, portfolio management, and regulatory compliance.

II. Materials and Methos

The data used in this study comprises borrower information from a publicly available dataset, including features such as age, income, employment status, credit history, loan amount, and payment behavior. Preprocessing steps involve handling missing values, normalizing numerical features, and encoding categorical variables. Feature selection techniques, such as correlation analysis and recursive feature elimination, are employed to retain only the most relevant variables.

A. Logistic Regression

Logistic regression serves as a baseline model. It predicts the probability of default using a logistic function, with coefficients estimated through maximum likelihood estimation. The model's simplicity and interpretability make it a valuable benchmark for comparing more complex algorithms. Here is the 1D figure illustrating the logistic regression model for credit risk assessment. It demonstrates the relationship between predictor variables and the probability of default, with a sigmoid curve representing the logistic function. The threshold line differentiates between lower and higher probabilities of default.

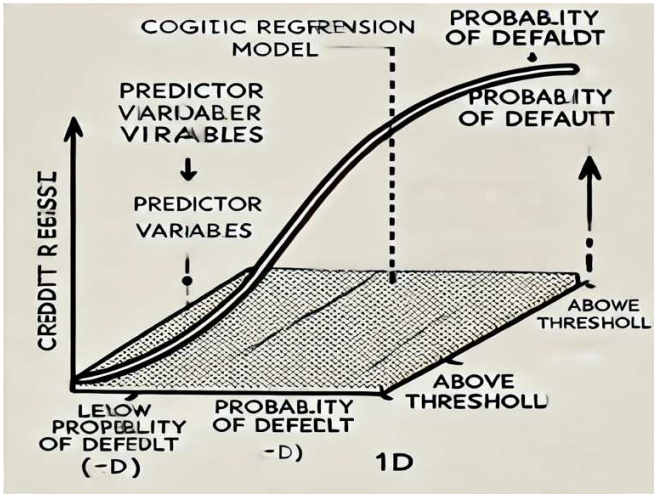

Fig. 1 Illustrating the logistic regression model for credit risk assessment

B. Decision Trees and Random Forest

Decision Trees are a popular machine learning technique used for classification and regression tasks. They work by recursively splitting a dataset into subsets based on feature values, creating a tree-like structure of decision nodes and leaf

nodes. Each node in the tree represents a decision based on a specific feature, with branches leading to child nodes that represent possible outcomes or further splits. The process continues until the tree reaches a leaf node, which represents a final decision or prediction. The primary advantage of decision trees is their interpretability, as they provide a clear and visual representation of decision-making processes. However, they can suffer from overfitting, especially when the tree is deep, meaning it has many levels, and perfectly fits the training data but fails to generalize well to unseen data.

Random Forests address the overfitting issue by combining multiple decision trees into an ensemble model. This ensemble approach involves training each tree on a different bootstrap sample of the data, which is a random sample with replacement. Additionally, at each split in the tree, Random Forests select a random subset of features to consider, further diversifying the model. The final prediction in a Random Forest is obtained by averaging the predictions of the individual trees in the ensemble for regression tasks or by taking the majority vote for classification tasks. This process of aggregation helps in reducing the variance of the model, thereby improving its generalization capability and robustness against overfitting.

The use of Random Forests has become widespread in various applications, including credit risk assessment, medical diagnosis, and environmental monitoring, due to their ability to handle large datasets with high-dimensional features and their resilience to noisy data. Moreover, Random Forests can provide insights into feature importance, as the model tracks how much each feature contributes to reducing the impurity in the trees. This capability is valuable for feature selection and understanding the underlying factors driving predictions.

In summary, Decision Trees and Random Forests are powerful tools in the machine learning toolbox, offering a balance of interpretability and predictive accuracy. Decision Trees provide a straightforward method for making decisions based on feature values, while Random Forests enhance this method by reducing overfitting and improving generalization through the ensemble of multiple trees. The combination of these methods makes them versatile and effective for a wide range of predictive modeling tasks.

C. Support Vector Machines (SVM)

Support Vector Machines (SVM) are a powerful and versatile class of supervised learning algorithms used for classification and regression tasks. In the context of credit risk assessment, SVMs [8], [9] classify borrowers by identifying the optimal hyperplane that distinctly separates two classes: defaulters and non-defaulters. The hyperplane is essentially a decision boundary that maximizes the margin, which is the distance between the hyperplane and the nearest data points from either class. These critical data points, known as support vectors, play a crucial role in defining the position and orientation of the hyperplane.

The strength of SVMs lies in their ability to handle both linear and non-linear relationships within data. While a linear SVM seeks to find a straight-line hyperplane in a high-dimensional space to separate the classes, not all data is linearly separable. To address this, SVMs utilize kernel functions to transform the original feature space into a higher-dimensional space. This transformation allows the algorithm to find a hyperplane in the transformed space, which corresponds to a non-linear decision boundary in the original space. Common kernel functions include the polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel [10], [11] each providing different ways of mapping the data.

Choosing the appropriate kernel and tuning regularization parameters are critical steps in optimizing an SVM model. The regularization parameter (often denoted as C) controls the trade-off between achieving a low error on the training data and maintaining a large margin to avoid overfitting. A high value of C prioritizes classifying all training examples correctly, which may result in overfitting, while a low value of C allows for a larger margin but may lead to misclassification. Cross-validation is typically used to fine-tune these parameters, ensuring that the model generalizes well to new, unseen data.

Beyond classification tasks, SVMs can also be adapted for regression, where the objective is to predict a continuous target variable. In this scenario, known as Support Vector Regression (SVR), the model aims to fit the data within a certain tolerance level, minimizing the margin of error while maintaining a margin around the hyperplane. SVMs are widely used in various domains, including finance, healthcare, and bioinformatics, due to their robustness and effectiveness in handling high-dimensional data. However, they can be computationally intensive, especially with large datasets, and the selection of the kernel function requires careful consideration and domain knowledge. Despite these challenges, the flexibility and precision of SVMs make them a valuable tool for predictive modeling, including credit risk assessment, where distinguishing between defaulters and non-defaulters is crucial for financial institutions.

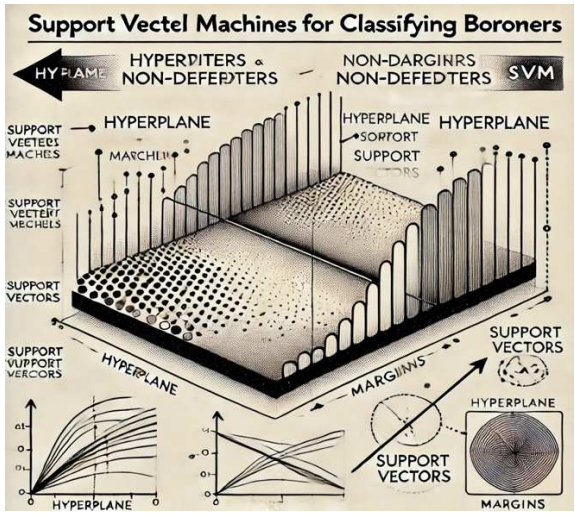

Fig. 2 Support Vectel Machines for Classifying Boroners

D. Neural Networks

Neural Networks are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected units called neurons, which are organized into layers. A feedforward neural network is a specific type of neural network where the data moves in one direction, from the input layer, through several hidden layers, to the output layer, without forming any cycles. This

architecture is particularly useful for tasks such as classification, regression, and pattern recognition.

The development of a feedforward neural network involves constructing multiple hidden layers, each comprising numerous neurons. These neurons are computational units that process input signals and generate output signals. The activation of each neuron is determined by an activation function, a non-linear mathematical function that introduces non-linearity into the model. Common activation functions include the Rectified Linear Unit (ReLU) [12] sigmoid, and hyperbolic tangent (tanh). The use of non-linear activation functions enables the network to learn complex relationships and patterns within the data, making it capable of handling a wide range of tasks.

Training a neural network involves adjusting the weights of the connections between neurons to minimize the error in the model's predictions. This is achieved through a process called backpropagation, which is an iterative optimization algorithm. During backpropagation, the network's predictions are compared to the actual outcomes using a loss function, which quantifies the difference between the predicted and actual values. The gradients of the loss function with respect to the weights are computed and used to update the weights in a direction that reduces the loss. This iterative process continues until the model converges to a set of weights that minimize the loss function, thereby improving the network's accuracy.

To prevent overfitting—a common problem where the neural network becomes too specialized in the training data and performs poorly on new, unseen data—several regularization techniques are employed. One such technique is dropout, where a random subset of neurons is temporarily deactivated during each training iteration. This prevents the network from becoming overly reliant on specific neurons and encourages a more robust learning process. Another regularization technique is L2 regularization (also known as weight decay) [13], [14] which adds a penalty to the loss function based on the magnitude of the weights. This discourages the model from assigning too much importance to any single feature, promoting generalization.

Neural networks have been widely adopted in various fields due to their flexibility and ability to model complex, non-linear relationships. They are used in applications ranging from image and speech recognition to natural language processing and financial forecasting. Despite their power, neural networks require careful tuning of hyperparameters, such as the number of layers, the number of neurons per layer, and the learning rate, to achieve optimal performance. Additionally, they can be computationally intensive and require substantial amounts of data for training. Nevertheless, the versatility and effectiveness of neural networks continue to drive advancements in artificial intelligence and machine learning, making them a cornerstone of modern data analysis and prediction.

E. Gradient Boosting Machines (GBM)

Gradient Boosting Machines (GBM) [15], [16] are a powerful ensemble learning technique used for both classification and regression tasks. The fundamental idea behind GBM is to build an ensemble of weak learners, which are simple models that perform only slightly better than random guessing. Typically, decision trees are used as the weak learners because they can easily handle complex relationships and interactions in the data. GBM constructs these trees in a sequential manner, where each subsequent tree aims to correct the errors made by the previous ones.

The process begins with the model training an initial tree based on the entire dataset, predicting the target variable. The differences between the actual and predicted values, known as residuals, are calculated. These residuals represent the errors made by the first tree. The next tree in the sequence is then trained to predict these residuals instead of the original target values. By focusing on the residuals, the second tree tries to capture the information that the first tree missed. This process is repeated, with each new tree being added to the ensemble to correct the errors made by the combined predictions of all previous trees.

One of the critical aspects of GBM is the use of a learning rate, which is a hyperparameter that controls the contribution of each tree to the final model. A smaller learning rate means that each tree has a smaller impact, and more trees are needed to fit the data. This can help in preventing overfitting, as it allows the model to make gradual improvements rather than large adjustments that might capture noise in the training data. The learning rate, along with other hyperparameters such as the depth of each tree and the number of trees, plays a crucial role in determining the model's performance.

The depth of the trees, also known as the maximum depth, controls how complex each tree can be. A deeper tree can capture more complex patterns but is also more prone to overfitting. Conversely, shallower trees are simpler and less likely to overfit, but they may not capture all the nuances in the data. The number of trees in the ensemble is another essential hyperparameter. While more trees can potentially lead to better performance, they also increase the computational cost and the risk of overfitting.

To fine-tune these hyperparameters, cross-validation is often used. This involves splitting the training data into multiple subsets and training the model on different combinations of these subsets. The model's performance is then evaluated on a validation set that was not used during training. This process helps in finding the optimal set of hyperparameters that balance the trade-off between underfitting and overfitting.

One of the advantages of GBM is its flexibility in handling different types of data and loss functions. For example, it can be used with different loss functions such as mean squared error for regression tasks or logistic loss for classification tasks. Additionally, GBM can handle missing data and is relatively robust to outliers.

In practical applications, GBM has been widely used in various domains, including finance, healthcare, and marketing. Its ability to model complex relationships and interactions in the data makes it particularly useful for predictive modeling. For instance, in credit scoring, GBM can accurately predict the likelihood of default by capturing intricate patterns in borrowers' financial histories and behaviors.

Despite its strengths, GBM also has some limitations. It can be computationally intensive, especially with a large number of trees or very deep trees. Additionally, it requires careful tuning of hyperparameters, and the training process can be

time-consuming. However, with the advent of efficient implementations and modern hardware, these challenges are becoming less of a barrier.

F. Model Evaluation

Models are evaluated using a variety of metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC-ROC) [17], [18]. Cross-validation is employed to ensure robustness, with the dataset divided into training and testing sets. The performance metrics guide the selection of the best model for predicting credit risk.

III. Result and Discussion

A. Model Performance

The models' performance is compared based on the evaluation metrics. Logistic regression, while simple and interpretable, shows limited accuracy due to its inability to capture non-linear relationships. Decision trees, though more flexible, suffer from overfitting, a problem mitigated by Random Forest and GBM [19], [20].

SVM demonstrates strong performance, particularly with a radial basis function (RBF) kernel [21], but at the cost of computational efficiency. Neural networks outperform other models in terms of accuracy and AUC-ROC, highlighting their capacity to learn complex patterns. However, their complexity and training time are notable drawbacks.

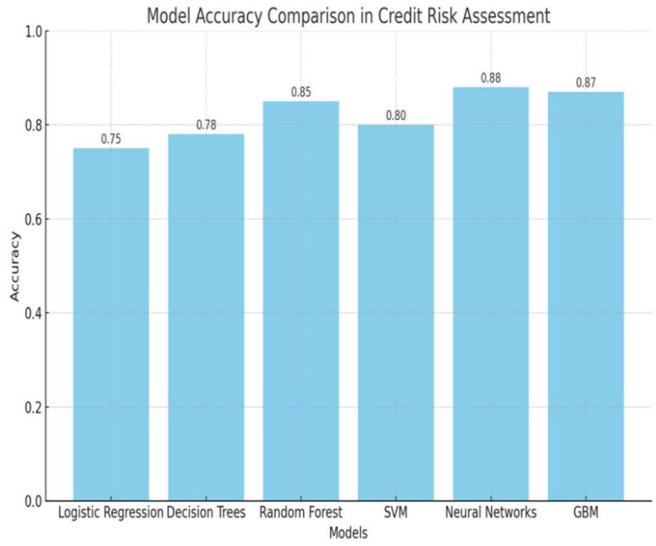

Fig. 3 Model Accuracy Comparison in Credit Risk Assessment

The bar chart illustrates the accuracy of various models in credit risk assessment, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), Neural Networks, and Gradient Boosting Machines (GBM). Among these, Neural Networks achieved the highest accuracy at 0.88, closely followed by GBM at 0.87. This indicates that these models are particularly effective at capturing complex patterns in the data, making them highly suitable for predicting credit risk. In contrast, Logistic Regression, while commonly used due to its simplicity and interpretability, has the lowest accuracy at 0.75, highlighting its limitations in handling non-linear relationships.

Decision Trees, with an accuracy of 0.78, show a moderate performance. Although they provide an intuitive way of understanding decision-making processes, their tendency to overfit the data can limit their predictive accuracy. This issue is mitigated in ensemble methods like Random Forest, which improves upon individual decision trees by averaging their predictions, resulting in a higher accuracy of 0.85. SVM also demonstrates robust performance with an accuracy of 0.80, benefiting from its capability to find optimal hyperplanes for classification in high-dimensional spaces.

The comparison underscores a key trade-off in model selection for credit risk assessment: while advanced models like Neural Networks and GBM offer superior accuracy, they often require more computational resources and data for training. Additionally, these models can act as "black boxes," making it challenging to interpret their predictions. On the other hand, simpler models like Logistic Regression and Decision Trees, though less accurate, provide greater transparency and ease of understanding. This highlights the importance of balancing predictive performance with interpretability, depending on the specific application and regulatory requirements in the financial sector.

IV. Conclusion

This paper provides a comprehensive analysis of various statistical and machine learning models used in credit risk assessment. The study compares traditional methods, such as logistic regression, with advanced machine learning techniques, including Decision Trees, Support Vector Machines (SVM), Neural Networks, and Gradient Boosting Machines (GBM). The findings indicate that Neural Networks and GBM offer superior predictive accuracy, effectively capturing complex, non-linear patterns in borrower data. These advanced models outperform traditional statistical methods, highlighting their potential in enhancing the accuracy and efficiency of credit risk assessments.

However, the study also underscores the importance of balancing model accuracy with interpretability, particularly in regulatory environments where transparency is crucial. While models like Neural Networks and GBM provide high accuracy, their "black box" nature can be a drawback in terms of understanding and explaining the decision-making process. In contrast, simpler models like logistic regression, despite their lower accuracy, offer greater transparency and ease of interpretation, making them valuable tools for certain applications.

The research also highlights the practical applications and challenges associated with these models. The need for substantial computational resources and large datasets for training advanced models poses challenges, particularly for smaller financial institutions. Moreover, the complexity of these models requires specialized expertise, which may not always be readily available.

In conclusion, the choice of a credit risk assessment model should be guided by the specific needs and constraints of the financial institution, including the trade-off between accuracy and interpretability. Future research should focus on developing hybrid models that combine the strengths of different approaches, potentially offering a balance between performance and transparency. Additionally, advancements in explainable AI could enhance the interpretability of

complex models, making them more accessible and usable in regulatory settings. This study provides a foundation for further exploration and development in the field of credit risk modeling, with the aim of optimizing decision-making processes and improving financial stability.

REFERENCES

[1] Md Abu Sufian Mozumder et al., "Revolutionizing Organizational Decision-Making for Banking Sector: A Machine Learning Approach with CNNs in Business Intelligence and Management," Journal of Business and Management Studies, vol. 6, no. 3, pp. 111–118, May 2024, doi: 10.32996/jbms.2024.6.3.12.

[2] Mohammad Anisur Rahman et al., "Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness," Journal of Business and Management Studies, vol. 6, no. 3, pp. 103–110, May 2024, doi: 10.32996/jbms.2024.6.3.11.

[3] Syeda Farjana Farabi et al., "Enhancing Credit Card Fraud Detection: A Comprehensive Study of Machine Learning Algorithms and Performance Evaluation," Journal of Business and Management Studies, vol. 6, no. 3, pp. 252–259, Jun. 2024, doi:10.32996/jbms.2024.6.13.21.

[4] R. Shahid et al., "Predicting Customer Sentiment in Social Media Interactions: Analyzing Amazon Help Twitter Conversations Using Machine Learning," International Journal of Advanced Science Computing and Engineering, vol. 6, no. 2, pp. 52–56, Jul. 2024, doi:10.62527/ijasce.6.2.211.

[5] R. Shahid et al., "Predicting Customer Loyalty in the Airline Industry: A Machine Learning Approach Integrating Sentiment Analysis and User Experience," International Journal on Computational Engineering, vol. 1, no. 2, pp. 50–54, Jun. 2024, doi:10.62527/comien.1.2.12.

[6] Chinmoy Modak et al., "Machine Learning Model in Digital Marketing Strategies for Customer Behavior: Harnessing CNNs for Enhanced Customer Satisfaction and Strategic Decision-Making," Journal of Economics, Finance and Accounting Studies, vol. 6, no. 3, pp. 178–186, Jun. 2024, doi: 10.32996/jefas.2024.6.3.14.

[7] Mohammad Anisur Rahman et al., "Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness," Journal of Business and Management Studies, vol. 6, no. 3, pp. 103–110, May 2024, doi: 10.32996/jbms.2024.6.3.11.

[8] Md Nasir Uddin Rana et al., "Revolutionizing Banking Decision-Making: A Deep Learning Approach to Predicting Customer Behavior," Journal of Business and Management Studies, vol. 6, no. 3, pp. 21–27, May 2024, doi: 10.32996/jbms.2024.6.3.3.

[9] MD Tanvir Islam et al., "Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce," Journal of Computer Science and Technology Studies, vol. 6, no. 1, pp. 33–39, Jan. 2024, doi:10.32996/jcsts.2024.6.1.4.

[10] Bishnu Padh Ghosh et al., "Deep Learning in Stock Market Forecasting: Comparative Analysis of Neural Network Architectures Across NSE and NYSE," Journal of Computer Science and Technology Studies, vol. 6, no. 1, pp. 68–75, Jan. 2024, doi:10.32996/jcsts.2024.6.1.8.

[11] Tuan Ngoc Nguyen et al., "Product Demand Forecasting with Neural Networks and Macroeconomic Indicators: A Comparative Study among Product Categories," Journal of Business and Management Studies, vol. 6, no. 2, pp. 170–175, Apr. 2024, doi:10.32996/jbms.2024.6.2.17.

[12] Mst Zannatun Ferdus, Nishat Anjum, Tuan Ngoc Nguyen, Afjal Hossain Jisan, and Md Azad Hossain Raju, "The Influence of Social Media on Stock Market: A Transformer-Based Stock Price Forecasting with External Factors," Journal of Computer Science and Technology Studies, vol. 6, no. 1, pp. 189–194, Feb. 2024, doi: 10.32996/jcsts.2024.6.1.20.

[13] R. Feldman & J. Sanger, Text mining handbook: Advanced approaches in analyzing unstructured data. New York, NY, USA: Cambridge University Press. 2006.

[14] Md Abu Sufian Mozumder et al., "Enhancing Customer Satisfaction Analysis Using Advanced Machine Learning Techniques in Fintech Industry," Journal of Computer Science and Technology Studies, vol. 6, no. 3, pp. 35–41, Aug. 2024, doi: 10.32996/jcsts.2024.6.3.4.

[15] Mohammad Shafiquzzaman Bhuiyan et al., "Advancements in Early Detection of Lung Cancer in Public Health: A Comprehensive Study Utilizing Machine Learning Algorithms and Predictive Models," Journal of Computer Science and Technology Studies, vol. 6, no. 1, pp. 113–121, Jan. 2024, doi: 10.32996/jcsts.2024.6.1.12.

[16] Md Tuhin Mia, Mst Zannatun Ferdus, Md Abdur Rakib Rahat, Nishat Anjum, Ummay Siddiqua, and Md Azad Hossain Raju, "A Comprehensive Review of Text Mining Approaches for Predicting Human Behavior using Deep Learning Method," Journal of Computer Science and Technology Studies, vol. 6, no. 1, pp. 170–178, Feb. 2024, doi: 10.32996/jcsts.2024.6.1.18.

[17] J. Miah and R. H. Khan, "Service Development of Smart Home Automation System," Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems, Nov. 2019, doi: 10.1145/3372422.3372437.

[18] R. H. Khan and J. Miah, "Performance Evaluation of a new one-time password (OTP) scheme using stochastic petri net (SPN)," 2022 IEEE World AI IoT Congress (AIIoT), Jun. 2022, doi:10.1109/aiiot54504.2022.9817203.

[19] R. H. Khan, J. Miah, S. M. Y. Arafat, M. M. M. Syeed, and D. M. Ca, "Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks," 2023 15th International Conference on Innovations in Information Technology (IIT), Nov. 2023, doi: 10.1109/iit59782.2023.10366426.

[20] R. H. Khan, J. Miah, M. M. M. Syeed, and M. F. Uddin, "Metaverse Ecosystem Realization for its Application Development," 2024 16th International Conference on Computer and Automation Engineering (ICCAE), Mar. 2024, doi: 10.1109/iccae59995.2024.10569237.

[21] 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.