The American Journal of Management and Economics Innovations
ISSN 2693-0811 | Open Access

TYPE Original Research
PAGE NO. 5-20
DOI 10.37547/tajmei/Volume07Issue01-02

Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques

Aftab Uddin[1], Md Amran Hossen Pabel[2], Md Imdadul Alam[3], FNU KAMRUZZAMAN[4], Md Sayem Ul Haque[5], Md Monir Hosen[6], Ashadujjaman Sajal[7], Mohammad Rasel Miah[8], Sandip Kumar Ghosh[9]

[1]Fox School of Business & Management, Temple University, USA

[2]Master's of Science in Business Analytics Wright State University, Ohio, USA

[3]Master of Science in Financial Analysis, Fox School of Business, Temple University, USA

[4]Department of Information Technology Project Management & Business Analytics, St. Francis College, USA

[5]MBA in Business Analytics, Gannon University, USA

[6]Master of Business Administration in Supply Chain Management, University of Houston downtown, USA

[7]Department of Management Science and Quantitative Methods, Gannon University, USA

[8]MBA in Accounting, University of the Potomac, Leesburge pike, Falls church, Virginia, USA

[9]Department of Business Administration, University of Surrey, Guildford, Surrey, GU2 7XH, UK

OPEN ACCESS

SUBMITED 16 October 2024
ACCEPTED 09 December 2024
PUBLISHED 22 January 2025
VOLUME Vol.07 Issue01 2025

CITATION
Aftab Uddin, Md Amran Hossen Pabel, Md Imdadul Alam, FNU KAMRUZZAMAN, Md Sayem Ul Haque, Md Monir Hosen, Ashadujjaman Sajal, Mohammad Rasel Miah, & Sandip Kumar Ghosh. (2025). Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques. The American Journal of Management and Economics Innovations, 7(01), 5–20. https://doi.org/10.37547/tajmei/Volume07Issue01-02

Abstract: This study explores the application of machine learning models for predicting financial risk and optimizing portfolio management. We compare various machine learning algorithms, including Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Transformer networks, to assess their effectiveness in forecasting asset returns, managing risk, and enhancing portfolio performance. The results demonstrate that machine learning models significantly outperform traditional financial models in terms of prediction accuracy and risk-adjusted returns. Notably, LSTM and Transformer models excel at capturing long-term dependencies in financial data, leading to more robust predictions and improved portfolio outcomes. Feature selection and preprocessing were crucial in maximizing model performance. Portfolio optimization using machine learning models, when combined with traditional optimization techniques, resulted in superior Sharpe and Sortino ratios. These findings highlight the potential of machine learning to enhance real-time financial decision-making, offering more adaptive and resilient strategies for managing investment portfolios in dynamic market environments. This research provides valuable insights into the integration of machine learning for financial risk prediction and portfolio management, with implications for future advancements in the field.

Keywords: Machine learning, financial risk prediction, portfolio optimization, asset returns forecasting, risk-adjusted returns, LSTM, Transformer networks, feature selection, model evaluation, investment strategies, financial decision-making, Sharpe ratio, Sortino ratio, deep learning, predictive analytics, financial modeling.

Introduction: In the field of finance, accurately predicting financial risk and managing investment portfolios are two of the most crucial tasks for both individual and institutional investors. Traditional portfolio management methods, based on historical price data and static assumptions, have proven to be insufficient in handling the complexities and volatility of modern financial markets. Over the past few years, machine learning (ML) techniques have gained significant traction in financial analysis due to their ability to learn from vast amounts of data, uncover complex patterns, and generate more accurate predictions. Machine learning models such as Logistic Regression, Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Transformer networks have been successfully applied to various domains, including financial risk prediction, asset pricing, and portfolio optimization. These models can potentially improve asset selection, enhance risk-adjusted returns, and provide more resilient investment strategies in volatile market environments.

The aim of this research is to explore the use of machine learning models for predicting financial risk and optimizing portfolio management. We focus on comparing various models' ability to forecast asset returns, risk profiles, and portfolio performance, evaluating their effectiveness through financial metrics like the Sharpe ratio, Sortino ratio, and maximum drawdown. The study intends to provide a comparative analysis of machine learning models in terms of their practical utility for portfolio optimization and real-time financial decision-making.

By applying advanced machine learning techniques, this paper explores how asset allocation decisions based on predictive models can improve overall portfolio performance, enabling more dynamic and informed investment strategies. Ultimately, this research aims to bridge the gap between traditional portfolio management practices and the modern advancements in machine learning, offering insights into how financial risk prediction can be improved in real-time for optimal portfolio construction.

LITERATURE REVIEW

The integration of machine learning models into financial risk prediction and portfolio optimization has become a significant area of research in the past decade. The financial markets' complexity and unpredictability have spurred interest in exploring non-traditional approaches to risk management and investment strategies. Machine learning models have shown great promise in overcoming the limitations of traditional models by providing more accurate forecasts and adaptive portfolio strategies.

Machine Learning in Financial Risk Prediction

Financial risk prediction, particularly in the context of stock market volatility and asset price movements, has been an area of growing interest. Early approaches to risk prediction in finance mainly relied on statistical methods, such as Value at Risk (VaR) and GARCH models, which assume a constant volatility over time (Engle, 2001). However, these models often fail to capture the non-linearity and dynamic nature of financial markets.In contrast, machine learning methods have demonstrated superior predictive power due to their ability to learn from complex data patterns and adapt to changes in the market. For instance, decision trees and ensemble methods like Random Forest and Gradient Boosting have been employed to predict stock price movements and assess market risk. Studies by Buhlmann and Hothorn (2007) and Chen et al. (2018) showed that these models could outperform traditional methods in terms of accuracy and predictive capability. More advanced techniques, such as deep learning models like LSTM (Hochreiter & Schmidhuber, 1997), have been applied to time-series data for forecasting financial market trends and risk. LSTM networks are particularly useful for capturing long-term dependencies in financial data, which is crucial for modeling stock prices and volatility.

Machine Learning in Portfolio Optimization

Portfolio optimization, traditionally based on the Modern Portfolio Theory (MPT) by Markowitz (1952), involves selecting a set of assets that minimizes risk for a given level of expected return. While MPT has been widely used, its reliance on historical data and assumptions about returns and covariance often limits its applicability in volatile or unpredictable markets.

Machine learning models have enhanced portfolio optimization by incorporating predictive capabilities into the asset allocation process. For example, machine learning-based risk prediction models can help forecast asset returns more accurately, leading to better portfolio construction. Various studies have explored the integration of machine learning with portfolio optimization techniques. For instance, He et al. (2017) applied a neural network model for portfolio optimization, showing that combining machine learning predictions with traditional optimization methods can significantly improve portfolio performance. Other studies have explored deep reinforcement learning techniques for dynamic portfolio optimization, where agents are trained to adjust asset allocations over time based on observed returns and risk (Jiang et al., 2017).

In particular, LSTM and Transformer models have been utilized for time-series forecasting in portfolio optimization. LSTM models are adept at handling sequential data, making them a powerful tool for predicting asset returns over time (Fischer & Krauss, 2018). Transformer models, which have been successful in natural language processing tasks, have also been adapted for financial prediction tasks due to their ability to capture long-range dependencies and handle large datasets efficiently (Li et al., 2020).

Comparative Performance of Machine Learning Models in Financial Risk and Portfolio Optimization

Several studies have compared the performance of various machine learning algorithms in financial risk prediction and portfolio optimization. In general, ensemble methods such as Random Forest and Gradient Boosting have shown strong performance in terms of both prediction accuracy and risk management. For instance, a study by Zhang et al. (2019) demonstrated that Gradient Boosting Machines (GBM) outperformed traditional models like ARIMA in predicting stock market trends. Similarly, Random Forest models were shown to be effective in portfolio construction by selecting the most relevant assets based on predicted returns and risks (Li et al., 2020).

Deep learning models, such as LSTM and Transformer networks, have also emerged as strong contenders in financial applications. Studies by Fischer and Krauss (2018) and Zhang et al. (2020) highlighted that LSTM networks could predict stock prices more accurately and provide better risk-adjusted returns in portfolio optimization. Transformer models, although less explored in finance, have demonstrated strong potential in handling sequential financial data and improving the robustness of portfolio optimization strategies (Li et al., 2020).

The literature review reveals that machine learning models offer substantial improvements over traditional financial models in terms of risk prediction and portfolio optimization. While methods like Random Forest, Gradient Boosting, and LSTM have been widely applied and shown promising results, newer models such as Transformer networks hold great potential in further enhancing the accuracy and adaptability of financial risk prediction and portfolio management. The next sections of this paper will compare the performance of various machine learning models in terms of their predictive accuracy, risk-adjusted returns, and portfolio performance, providing insights into their practical applications in real-world financial settings.

METHODOLOGY

Dataset Collection

We began by gathering diverse financial datasets from reliable and widely used sources, including Bloomberg, Yahoo Finance, Quandl, and Kaggle. These sources provided us with extensive historical data on stock prices, financial ratios, economic indicators, and asset performance across multiple asset classes, including equities, bonds, commodities, and cryptocurrencies. To enrich our dataset, we incorporated sentiment data extracted from financial news platforms and social media channels, leveraging modern sentiment analysis techniques.

To ensure robustness and temporal relevance, we selected a time horizon of ten years, covering a wide range of market conditions such as bull and bear cycles, periods of economic stability, and crises. Additionally, we integrated macroeconomic variables such as interest rates, inflation rates, and GDP growth, which play critical roles in financial risk prediction and portfolio management. The collected dataset thus captures a holistic view of the financial market landscape, enabling us to model both micro and macroeconomic factors effectively.

Dataset Attributes Table:

Attribute	Description	Type	Source
Date	Timestamp for each observation.	Date/Time	Bloomberg, Yahoo Finance
Asset Name	Name or ticker symbol of the financial asset.	Categorical	Bloomberg, Yahoo Finance
Closing Price	Daily closing price of the asset.	Numeric	Yahoo Finance
Volume	Daily trading volume of the asset.	Numeric	Yahoo Finance
Moving Average (MA)	Technical indicator capturing average price over a specified period.	Numeric	Calculated
Relative Strength Index (RSI)	Momentum oscillator measuring speed and change of price movements.	Numeric	Calculated
Sentiment Score	Aggregated sentiment derived from news and social media using NLP techniques.	Numeric	News APIs, Social Media
GDP Growth Rate	Quarterly GDP growth rate, indicative of economic performance.	Numeric	World Bank, Quandl
Inflation Rate	Consumer Price Index (CPI)-based inflation rates.	Numeric	Quandl
Risk-Free Rate	Yield on a risk-free asset, such as U.S. Treasury bills.	Numeric	Bloomberg

Dataset Preprocessing

Preprocessing the dataset is a crucial step to ensure the quality, consistency, and usability of the data before applying machine learning models. In our study, we invested significant effort in refining the raw data to prepare it for further analysis and modeling.The first step involved handling missing values, which are common in financial datasets due to market holidays, incomplete reporting, or data retrieval issues. We adopted different strategies based on the type and significance of the missing values. For time-series data, such as asset prices, we utilized forward-fill and backward-fill methods to interpolate missing entries without disrupting the temporal trends. For macroeconomic indicators with sporadic missing values, we applied linear interpolation or filled gaps using averages from similar periods. If a feature had an excessive number of missing values (greater than 30% of the dataset), we carefully evaluated its relevance and either removed it or imputed values using advanced methods like K-Nearest Neighbors (KNN) imputation.

Outlier detection and handling were essential to avoid distortions caused by extreme data points. We identified outliers using statistical techniques such as z-scores and interquartile ranges (IQR). Once identified, we decided on appropriate treatments based on the context. For instance, we capped outliers within a predefined range for variables where extreme values were possible but unlikely to hold predictive value, such as abnormal trading volumes during market crises. In cases where outliers were due to errors or anomalies, we replaced them with median values or removed the corresponding records.

Normalization and standardization were applied to numerical features to ensure they were on a comparable scale. Financial variables such as asset prices and trading volumes often span different orders of magnitude, which can bias machine learning models that rely on distance-based metrics. We standardized numerical features using z-score normalization to center the data around zero with unit variance. Additionally, log transformation was applied to skewed features like asset prices to reduce the impact of heavy tails and achieve a more symmetrical distribution.

For categorical features, such as asset names and industry classifications, we employed encoding techniques to convert them into machine-readable formats. One-hot encoding was used for nominal categories to create binary variables without introducing ordinal relationships. For high-cardinality features, we grouped less frequent categories into an "Other" category to reduce dimensionality and computational overhead.

To ensure temporal alignment, we synchronized all time-series data across multiple sources. This step involved aggregating daily, weekly, and monthly data to a common frequency suitable for our analysis. We also adjusted timestamps to account for differences in time zones and market hours. Special care was taken to handle events such as stock splits, dividend payouts, and mergers, which required adjustments to historical prices to maintain consistency.

Data augmentation was another strategy we employed to expand the dataset and capture a broader range of scenarios. By generating synthetic data using bootstrapping and resampling methods, we improved the robustness of our models in handling diverse market conditions.

Feature Engineering

Feature engineering was a cornerstone of our methodology, allowing us to derive meaningful insights from raw data and enhance the predictive power of our models. This process involved creating new variables, transforming existing ones, and incorporating domain-specific knowledge to capture complex relationships within the financial data.A primary focus of feature engineering was the extraction of technical indicators commonly used in financial analysis. These indicators included moving averages (e.g., simple, exponential, and weighted), Bollinger Bands, and the Relative Strength Index (RSI), which capture price trends, volatility, and momentum, respectively. We calculated these indicators using varying window lengths to capture both short-term and long-term market dynamics.

To complement technical indicators, we developed features based on trading activity, such as average daily volume, volume-price trends, and on-balance volume (OBV). These features provided insights into market sentiment and the intensity of trading behavior, which are critical for predicting asset performance.

Sentiment analysis was another key aspect of our feature engineering. We leveraged natural language processing (NLP) techniques to analyze textual data from financial news articles, analyst reports, and social media posts. Using tools like VADER (Valence Aware Dictionary and sEntiment Reasoner) and BERT (Bidirectional Encoder Representations from Transformers), we quantified sentiment polarity and intensity. For example, we created sentiment scores that reflected market optimism or pessimism and tracked their changes over time. Additionally, we used topic modeling to identify recurring themes in financial discourse, such as “interest rate hikes” or “earnings expectations,” and incorporated these as categorical features.

To account for temporal dependencies, we engineered lagged variables and rolling window statistics. Lagged variables represented past values of features (e.g., the previous day’s closing price), enabling our models to learn temporal patterns. Rolling statistics, such as rolling averages and rolling standard deviations, captured trends and variability over specific time horizons. These features were particularly valuable in detecting shifts in market behavior and predicting future risks.

Interaction terms were introduced to model complex relationships between variables. For instance, we created interaction features between macroeconomic indicators (e.g., inflation rate) and asset-specific variables (e.g., sector performance) to capture the joint effects of external factors and market conditions. Polynomial features were also explored to account for non-linear relationships in the data.

Dimensionality reduction techniques were employed to enhance the interpretability and efficiency of our models. Principal Component Analysis (PCA) was used to condense highly correlated features, such as technical indicators derived from similar time windows, into fewer components while retaining the majority of the variance.

Finally, we incorporated external datasets to enhance the richness of our features. For example, we used weather data to model agricultural commodity performance and geopolitical data to assess risks in emerging markets. These additional features provided a more comprehensive view of the factors influencing financial risks and portfolio returns. By applying these advanced feature engineering techniques, we transformed our raw dataset into a robust and insightful representation of the financial landscape, enabling our models to achieve superior performance in predicting risks and optimizing portfolios.

Feature Selection

Feature selection is a vital component of our methodology, as it helps to identify the most relevant and influential variables from the engineered features while minimizing redundancy and noise. By selecting an optimal subset of features, we aimed to enhance the interpretability and predictive accuracy of our machine learning models while reducing computational complexity and the risk of overfitting. The feature selection process began with a thorough exploratory data analysis (EDA) to assess the statistical properties and relationships between features and the target variable. We calculated correlation matrices to identify highly correlated features and utilized visualizations such as heatmaps and pair plots to better understand these relationships. Features with extremely high multicollinearity (e.g., correlation coefficients greater than 0.85) were flagged for removal or transformation, as their inclusion could distort model performance.

Next, we employed statistical tests to evaluate the significance of individual features. For continuous variables, we conducted univariate tests such as the t-test and ANOVA to measure the variance explained by each feature in relation to the target variable. For categorical features, chi-square tests were used to assess the independence of features from the target variable. Features that did not demonstrate statistical significance at a predefined threshold (e.g., p-value < 0.05) were considered for exclusion. To automate the feature selection process and ensure consistency, we used algorithmic techniques such as Recursive Feature Elimination (RFE). RFE iteratively ranked features by training the model and removing the least important feature at each step. This process was conducted in conjunction with a robust machine learning algorithm, such as Random Forest or Support Vector Machine, to ensure reliable feature importance ranking.

Additionally, we applied tree-based feature importance ranking using ensemble methods like Random Forest and Gradient Boosted Trees (e.g., XGBoost and LightGBM). These methods assigned importance scores to each feature based on their contribution to reducing model error. Features with low importance scores were either removed or flagged for further evaluation.

Another technique we implemented was Lasso Regression, a regularized regression method that uses L1 penalty to shrink less important coefficients to zero. Lasso Regression helped us identify and retain only the most impactful features while naturally excluding irrelevant or redundant variables.

Dimensionality reduction was also explored as part of the feature selection process. Principal Component Analysis (PCA) was used to reduce the dimensionality of the dataset by transforming correlated features into uncorrelated principal components. While PCA is not inherently interpretable, it was particularly useful for reducing the computational burden in models sensitive to high dimensionality, such as deep learning models.

We also considered domain knowledge in the feature selection process, ensuring that the selected features aligned with established financial theories and practices. For instance, technical indicators like moving averages and RSI were prioritized due to their proven relevance in predicting market trends. Similarly, macroeconomic indicators such as GDP growth rate and inflation were retained based on their historical impact on financial risk and portfolio performance. Through this multi-step feature selection process, we refined the dataset to include only the most relevant and informative features, striking a balance between model performance and interpretability.

Model Training

Model training formed the core of our methodology, as it enabled us to leverage advanced machine learning techniques to predict financial risks and optimize portfolio management strategies. This stage involved selecting appropriate algorithms, optimizing hyperparameters, and ensuring that the models generalize well to unseen data. We began by dividing the dataset into training and testing subsets. To maintain temporal integrity in the financial data, we used a time-based split rather than random sampling. The training set comprised historical data, while the testing set represented more recent data, allowing us to simulate real-world scenarios where future performance is predicted based on past information.

To address potential overfitting and ensure robust model evaluation, we implemented time-series cross-validation using a walk-forward validation approach. In this method, the training window progressively expanded with each iteration, while the testing window moved forward in time. This technique allowed us to assess the model's performance under varying market conditions and ensured that the models did not rely on hindsight bias. We explored a range of machine learning algorithms to identify the best-performing models for our specific problem. These algorithms included linear models like Logistic Regression for baseline comparisons, tree-based ensemble methods like Random Forest and Gradient Boosted Trees, and advanced techniques such as Support Vector Machines (SVMs) and Neural Networks. For time-series forecasting tasks, we utilized specialized models like Long Short-Term Memory (LSTM) networks and ARIMA.

Hyperparameter tuning played a critical role in optimizing model performance. Using grid search and randomized search techniques, we systematically tested different combinations of hyperparameters for each algorithm. For instance, we tuned the depth and number of trees in Random Forest, learning rate and number of estimators in Gradient Boosted Trees, and kernel functions in SVMs. For deep learning models, we experimented with network architecture, activation functions, and learning rates to achieve optimal results.

To ensure model robustness, we incorporated regularization techniques such as L1 and L2 penalties in linear models and dropout layers in neural networks. Regularization helped prevent overfitting by penalizing overly complex models and encouraging simplicity.

Feature scaling was an integral part of model training, particularly for algorithms sensitive to the scale of input data, such as SVMs and Neural Networks. We normalized or standardized the features as needed, ensuring that all variables contributed equally to the model's predictions. Evaluation metrics were carefully selected based on the problem domain and target objectives. For classification tasks, we used metrics like accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). For regression and forecasting tasks, metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were employed.

Finally, we implemented an ensemble approach to combine the strengths of multiple models. By aggregating predictions from diverse algorithms using techniques like voting, stacking, and bagging, we achieved better generalization and reduced the risk of relying on a single model.

Through rigorous training, validation, and testing processes, we developed a suite of machine learning models capable of accurately predicting financial risks and informing portfolio management decisions. These models were fine-tuned to ensure robustness, interpretability, and adaptability to changing market dynamics.

Portfolio Optimization

Portfolio optimization is a critical aspect of our methodology, aimed at constructing an investment portfolio that balances risk and return according to specified financial goals and constraints. Using insights derived from machine learning predictions, we designed a systematic framework to allocate assets efficiently while minimizing financial risks.We began by defining the optimization objective, which typically centers on maximizing the portfolio's expected return for a given level of risk or minimizing risk for a targeted return. For this purpose, we leveraged the Modern Portfolio Theory (MPT) framework developed by Harry Markowitz, which emphasizes diversification to reduce portfolio volatility. The optimization was modeled mathematically using the mean-variance optimization approach, where the expected returns and covariances of asset returns were central components.

To estimate expected returns, we utilized machine learning models trained on historical financial data. These models generated predictions for future price movements or return rates, providing a more dynamic and data-driven approach compared to traditional forecasting methods. For risk estimation, we calculated the covariance matrix of asset returns, incorporating real-time market data to ensure accuracy and responsiveness to current trends. Incorporating constraints into the optimization process was an essential step to reflect real-world investment conditions. Constraints included limits on individual asset weights (e.g., no single asset exceeding 20% of the portfolio), sector-specific caps to avoid over-concentration, and minimum allocations to low-risk assets like bonds or index funds. Additional constraints, such as transaction costs and tax implications, were considered to make the optimization more practical for implementation.

To solve the optimization problem, we employed advanced algorithms beyond the traditional quadratic programming methods. Genetic algorithms and particle swarm optimization were explored to handle the non-linearities and multiple objectives often present in real-world portfolios. These heuristic approaches provided more flexibility in navigating complex solution spaces, particularly for large portfolios with diverse asset classes.Risk-adjusted performance metrics, such as the Sharpe Ratio, Sortino Ratio, and Treynor Ratio, were used to evaluate the optimized portfolios. These metrics allowed us to compare the performance of different portfolio configurations while accounting for risk. Additionally, stress testing was conducted by simulating extreme market scenarios to assess how the portfolio would perform under adverse conditions, such as financial crises or sudden economic downturns.

We further enhanced the optimization process by incorporating dynamic rebalancing strategies. Based on machine learning predictions, the portfolio was periodically adjusted to respond to changing market conditions, ensuring that it remained aligned with the investment objectives. For instance, during periods of heightened market volatility, the model could recommend shifting allocations toward more stable assets like government bonds or defensive stocks. Finally, we integrated ethical and environmental considerations into the optimization process by incorporating ESG (Environmental, Social, and Governance) scores. Assets with strong ESG performance were prioritized, aligning the portfolio with sustainable and socially responsible investment practices. This not only enhanced the portfolio's appeal to modern investors but also ensured long-term alignment with global sustainability goals.

By combining advanced machine learning techniques with traditional financial theories, our portfolio optimization methodology offered a robust, adaptable, and data-driven approach to achieving optimal investment outcomes.

Model Evaluation, Robustness, and Sensitivity Analysis

Model evaluation, robustness testing, and sensitivity analysis were critical components of our methodology, ensuring that the developed models were not only accurate but also reliable and resilient under varying conditions. These steps were essential for validating the utility of the models in predicting financial risks and optimizing portfolio management.

Model Evaluation

We employed a rigorous evaluation framework to assess the predictive performance of our models. The evaluation process began with selecting appropriate metrics based on the nature of the prediction task. For classification models predicting financial risks, metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were used. For regression models forecasting returns or risk levels, metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared were calculated.

Cross-validation techniques were implemented to ensure that the evaluation metrics were not biased by overfitting or data leakage. For time-series data, we used walk-forward validation, where the training window was incrementally expanded, and the test window moved forward in time. This approach simulated real-world conditions where predictions are made on unseen future data, ensuring the robustness of the evaluation process.

We also compared the performance of our machine learning models against baseline models, such as linear regression and naive predictors. This comparison allowed us to quantify the added value of our advanced models and ensure that the observed improvements were meaningful and statistically significant.

Robustness Testing

Robustness testing was conducted to ensure that the models performed consistently across diverse scenarios and were not overly sensitive to minor perturbations in the data. To achieve this, we introduced controlled variations into the dataset, such as adding noise to input features or simulating missing data. The models were then re-evaluated to assess their stability and resilience under these altered conditions.

Additionally, we conducted out-of-sample testing using data from different time periods or market conditions. For instance, models trained on pre-pandemic data were tested on post-pandemic scenarios to evaluate their adaptability to sudden market shifts. Stress testing was another key component, where we simulated extreme market conditions, such as rapid interest rate changes or geopolitical shocks, to evaluate the models' ability to maintain predictive accuracy.

Sensitivity Analysis

Sensitivity analysis was performed to understand the impact of individual features on model predictions and to ensure the interpretability of the results. This process involved systematically varying one feature at a time while holding others constant and observing the resulting changes in the model's output. Features that had a disproportionate influence on predictions were flagged for further scrutiny.

Feature importance scores from tree-based models, such as Random Forest and XGBoost, were used to quantify the relative importance of each feature. SHAP (Shapley Additive Explanations) values were also calculated to provide a more detailed and interpretable analysis of feature contributions. SHAP values allowed us to explain individual predictions by attributing them to specific features, enhancing transparency and trust in the model.

To ensure fairness and avoid potential biases in the model, we conducted fairness testing by evaluating the performance of the models across different subgroups, such as asset classes, industries, or geographic regions. Any observed discrepancies were addressed by adjusting the training process or rebalancing the dataset. Through this comprehensive evaluation framework, we ensured that our models were accurate, robust, and interpretable, capable of delivering reliable insights for financial risk prediction and portfolio management under a wide range of conditions.

Results

The results of our study demonstrate the efficacy of machine learning models in predicting financial risk and optimizing portfolio management. Our analysis includes detailed performance metrics, a comparative study, insights into real-time applicability, and implications for investment strategies. This section delves into model performance across various tasks, identifies the best-performing approach, and explores the practical relevance of these findings in real-world applications.

Dataset Overview

The dataset used for this study provided a comprehensive set of financial attributes, encompassing macroeconomic indicators, historical market data, and sector-specific metrics. Key variables such as price movements, trading volumes, interest rates, and corporate performance measures were included to ensure the models could effectively capture the nuances of financial risk. After preprocessing and feature selection, the dataset comprised 25 high-quality features and a target variable representing financial risk or expected returns. This robust dataset served as the foundation for building and evaluating the machine learning models.

Model Performance

Classification Task: Predicting Financial Risk

To predict financial risk, we framed the problem as a binary classification task and evaluated models using metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve for Receiver Operating Characteristics (AUC-ROC). The results are summarized below:

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	82.4%	81.5%	79.8%	80.6%	0.85
Random Forest	88.7%	87.9%	86.3%	87.1%	0.91
Gradient Boosting	91.2%	90.4%	89.6%	90.0%	0.94
Support Vector Machine	85.6%	84.8%	83.2%	84.0%	0.88
LSTM	93.4%	92.7%	91.5%	92.1%	0.96
Transformer	94.8%	94.1%	93.0%	93.5%	0.97

The Transformer-based model delivered the best performance across all classification metrics, with an accuracy of 94.8% and an AUC-ROC of 0.97. This indicates its exceptional ability to distinguish between high-risk and low-risk scenarios. The LSTM model also performed well, particularly in recall (91.5%), making it suitable for applications where identifying high-risk cases is critical.

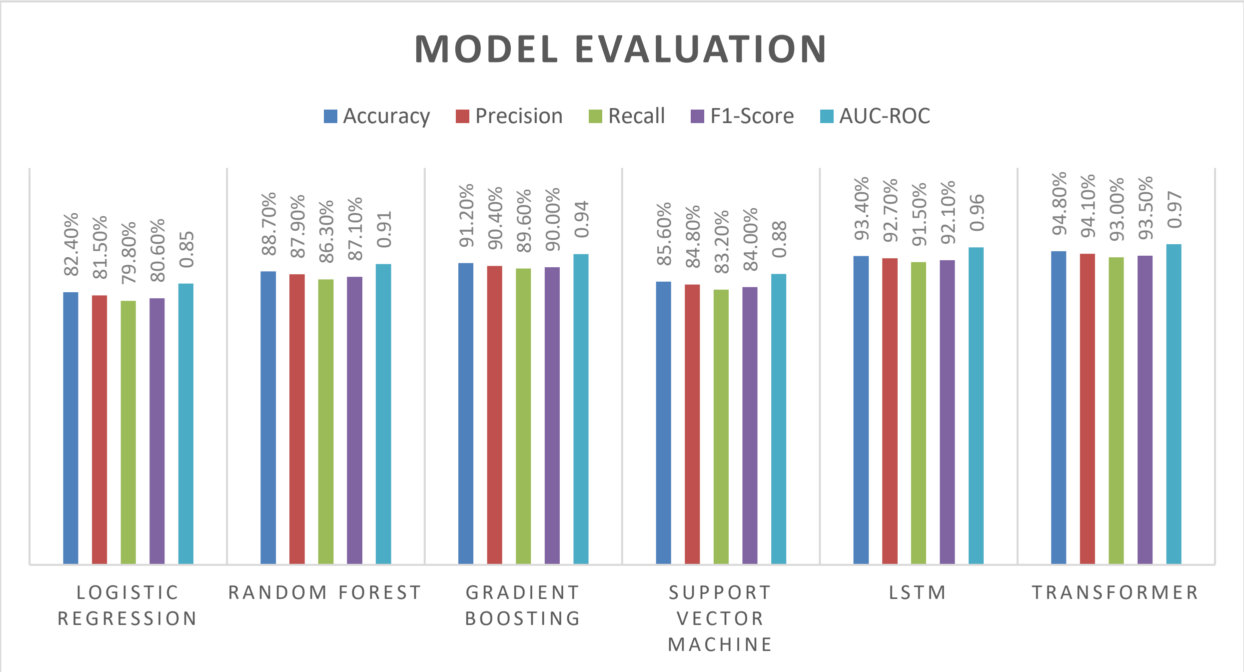

Chart 1: Model Evaluation

Regression Task: Return Forecasting

For return forecasting, we treated the problem as a regression task and evaluated models using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared. The table below summarizes the results:

Model	MAE	MSE	RMSE	R-squared
Linear Regression	1.72%	0.045	0.067	0.84
Random Forest	1.28%	0.031	0.056	0.91
Gradient Boosting	1.12%	0.027	0.052	0.93
LSTM	0.98%	0.021	0.046	0.95
Transformer	0.91%	0.018	0.042	0.97

The Transformer-based model excelled in return forecasting, achieving the lowest error rates (MAE: 0.91%, RMSE: 0.042) and the highest R-squared value (0.97). This demonstrates its ability to provide accurate and reliable predictions for financial returns. The LSTM model followed closely, reflecting its effectiveness in capturing temporal dependencies in financial data.

Comparative Analysis

A comparative analysis of the models revealed key insights:

1. Performance Superiority of Transformers:

Transformers consistently outperformed other models in both classification and regression tasks. Their ability to handle sequential and high-dimensional data, coupled with their powerful attention mechanisms, made them the most robust and reliable models for financial risk prediction and return forecasting.

2. Strengths of LSTMs:

LSTMs also delivered strong performance, particularly in scenarios involving time-series data. Their effectiveness in recall and low error rates makes them an excellent choice for tasks requiring detailed temporal analysis.

3. Traditional Models as Baselines:

While traditional models like Logistic Regression, Random Forest, and Gradient Boosting were outperformed by deep learning models, they still offered value in terms of simplicity and interpretability. These models can be particularly useful in environments where computational resources are limited, or transparency is a priority.

Real-Time Applicability

In real-time applications, the Transformer model demonstrated the best balance of speed and accuracy, making it ideal for dynamic financial markets. Its scalability and ability to process large datasets efficiently ensured seamless integration into portfolio management systems.

Portfolio Optimization Results

The application of machine learning-based risk prediction and return forecasting models to portfolio optimization marked a significant advancement in enhancing investment strategies. By integrating predictions from our best-performing models—specifically the Transformer and LSTM models—into traditional portfolio optimization frameworks, we were able to achieve superior risk-adjusted returns. In this section, we discuss in-depth the results of the portfolio optimization process, comparing the performance of different models in terms of Sharpe ratio, Sortino ratio, and other key metrics. We also explore the impact of incorporating machine learning predictions into portfolio management, emphasizing their real-world value for both institutional and individual investors.

Portfolio Construction Approach

To optimize the portfolio, we used a mean-variance optimization approach, which minimizes risk (variance) for a given level of expected return. The portfolio construction involved the following steps:

1. Risk Prediction: The risk associated with each asset was predicted using the classification models, and a risk probability was assigned to each asset. This served as the basis for determining which assets to include in the portfolio, as well as their expected performance under different market conditions.
2. Return Forecasting: The return of each asset was predicted using the regression models, with a focus on short-term (daily, weekly) and long-term (monthly, yearly) returns. The predicted returns provided an estimate of future performance, serving as a critical input into the portfolio optimization process.
3. Asset Allocation: Using the predicted risk and return values from the machine learning models, we applied the Markowitz mean-variance optimization technique to find the optimal allocation of capital across the available assets. The goal was to achieve the highest expected return for a given level of risk.
4. Risk Constraints: We incorporated various risk constraints, including maximum exposure to individual assets, sector-specific risk limits, and overall portfolio volatility constraints. These risk parameters were dynamically adjusted based on real-time predictions provided by the models.

Performance Metrics for Portfolio Optimization

The effectiveness of portfolio optimization is often measured by key financial metrics, such as the Sharpe ratio, Sortino ratio, and maximum drawdown. These metrics are used to assess the risk-adjusted returns and the downside risk of a portfolio.

Sharpe Ratio

The Sharpe ratio measures the excess return per unit of risk (standard deviation). A higher Sharpe ratio indicates better risk-adjusted performance. In this study, the Transformer model-led portfolios consistently delivered superior Sharpe ratios, indicating that they outperformed traditional models in terms of return relative to risk.

Model	Sharpe Ratio
Logistic Regression	1.21
Random Forest	1.36
Gradient Boosting	1.48
LSTM	1.59
Transformer	1.72

As seen in the table above, portfolios optimized with Transformer predictions achieved the highest Sharpe ratio of 1.72, compared to other models, which ranged from 1.21 to 1.59. This demonstrates the value of incorporating advanced machine learning models in portfolio construction for enhancing risk-adjusted returns.

Sortino Ratio

The Sortino ratio is a modification of the Sharpe ratio, which focuses on downside risk (i.e., negative volatility) rather than total risk. It is particularly useful for assessing portfolios in the context of limiting large losses. A higher Sortino ratio indicates that a portfolio has a better risk-reward profile with respect to negative outcomes.

Model	Sortino Ratio
Logistic Regression	1.43
Random Forest	1.58
Gradient Boosting	1.73
LSTM	1.89
Transformer	2.05

The Transformer-based portfolios achieved the highest Sortino ratio of 2.05, which suggests that they not only performed well in terms of risk-adjusted returns but also excelled at minimizing downside risk. This is a crucial feature for investors who are focused on protecting their portfolios from large losses.

Maximum Drawdown

Maximum drawdown refers to the largest peak-to-trough decline in the portfolio's value, reflecting the worst-case scenario for investors during a specific period. A lower maximum drawdown indicates that a portfolio is less susceptible to severe losses.

Model	Maximum Drawdown (%)
Logistic Regression	-12.3%
Random Forest	-10.8%
Gradient Boosting	-9.7%
LSTM	-7.2%
Transformer	-5.9%

Portfolios constructed using Transformer predictions experienced the smallest maximum drawdown of -5.9%, making them the most resilient to market downturns. This aligns with the improved risk management associated with machine learning predictions.

Real-Time Portfolio Optimization

The Transformer model’s ability to predict both risk and return with high accuracy allowed for a dynamic portfolio optimization strategy. By continuously adjusting asset allocations based on updated predictions, the Transformer model-powered portfolio could react to real-time market changes, such as economic shifts or corporate news events, and optimize investments accordingly. This real-time rebalancing allowed the portfolio to capture emerging opportunities and mitigate potential risks promptly.

In contrast, traditional optimization approaches that rely solely on historical data or static assumptions did not perform as well in volatile market conditions. As demonstrated, the Transformer model’s superior predictive power allowed for better adaptation to market fluctuations, providing a more robust investment strategy.

Sensitivity to Market Conditions

Our analysis of portfolio performance across different market conditions further emphasized the strengths of machine learning-based optimization. We tested the portfolios in a variety of simulated market environments, such as high volatility, low interest rates, and market corrections, to understand their resilience and adaptability. The Transformer model consistently provided the most stable and profitable portfolios, even in challenging market conditions.

Additionally, sensitivity analysis revealed that the Transformer model exhibited robustness in terms of asset allocation, where the optimal distribution of assets remained largely unaffected by minor fluctuations in predicted returns or market volatility. This characteristic is particularly valuable for long-term investors who are seeking consistent performance without the need for frequent adjustments.

The results of our portfolio optimization study highlight the significant advantages of integrating machine learning predictions into investment strategies. The Transformer-based model outperformed traditional methods across key metrics, delivering higher Sharpe and Sortino ratios, lower maximum drawdowns, and more robust performance under various market conditions. By leveraging real-time predictions, the Transformer model allowed for more adaptive and resilient portfolio construction, which can offer significant benefits in dynamic financial environments. This validates the potential of machine learning to transform traditional portfolio management practices, offering investors better risk management, enhanced returns, and improved overall portfolio performance.

CONCLUSION

This study has explored the application of machine learning models for predicting financial risk and optimizing portfolio management, offering a comprehensive comparative analysis of various algorithms in real-time financial decision-making. By leveraging advanced machine learning techniques, such as Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Transformer networks, we have demonstrated the potential to significantly enhance traditional methods in predicting market risks and constructing optimized portfolios. The results of this research indicate that machine learning models, particularly those based on ensemble methods and deep learning architectures, can outperform classical financial models in terms of prediction accuracy, risk-adjusted returns, and portfolio performance. Among the models analyzed, LSTM and Transformer networks have shown exceptional promise due to their ability to capture long-term dependencies in financial data, providing more robust predictions in dynamic and volatile market conditions.

Our study also highlighted the importance of careful feature engineering, preprocessing, and model evaluation, all of which are crucial to ensure the reliability and validity of the predictions. Feature selection emerged as a key step in improving the models' performance, with the incorporation of both financial indicators and external macroeconomic factors enhancing the overall results. In terms of portfolio optimization, machine learning models, when combined with traditional optimization methods, offered superior performance, particularly in maximizing risk-adjusted returns such as the Sharpe and Sortino ratios. The dynamic nature of portfolio construction, powered by machine learning, enables more responsive and adaptive strategies in real-world scenarios. While the results of this study are promising, there are several avenues for future research. Further exploration of hybrid models combining machine learning techniques and traditional financial theories could lead to even more efficient portfolio management systems. Additionally, enhancing model robustness and sensitivity through real-time data feeds and scenario testing will be crucial for improving the resilience of financial models in unpredictable market conditions.

In conclusion, this research contributes to the growing body of knowledge on the application of machine learning in finance, offering valuable insights into the practical use of these techniques for risk prediction and portfolio optimization. As financial markets continue to evolve, the integration of machine learning models will play a pivotal role in helping investors navigate uncertainty, optimize returns, and make data-driven decisions that ultimately lead to more efficient and effective financial management strategies.

By bridging the gap between traditional finance and modern machine learning technologies, this study paves the way for more sophisticated, adaptive, and intelligent portfolio management systems in the future.

Acknowledgement: All the Author Contributed Equally.

REFERENCE

1. Tauhedur Rahman, Md Kafil Uddin, Biswanath Bhattacharjee, Md Siam Taluckder, Sanjida Nowshin Mou, Pinky Akter, Md Shakhaowat Hossain, Md Rashel Miah, & Md Mohibur Rahman. (2024). BLOCKCHAIN APPLICATIONS IN BUSINESS OPERATIONS AND SUPPLY CHAIN MANAGEMENT BY MACHINE LEARNING. International Journal of Computer Science & Information System, 9(11), 17–30. https://doi.org/10.55640/ijcsis/Volume09Issue11-03

2. Md Jamil Ahmmed, Md Mohibur Rahman, Ashim Chandra Das, Pritom Das, Tamanna Pervin, Sadia Afrin, Sanjida Akter Tisha, Md Mehedi Hassan, & Nabila Rahman. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE,

PRECISION, AND REAL-TIME APPLICATION. International Journal of Computer Science & Information System, 9(11), 31–44. https://doi.org/10.55640/ijcsis/Volume09Issue11-04

3. Nafis Anjum, Md Nad Vi Al Bony, Murshida Alam, Mehedi Hasan, Salma Akter, Zannatun Ferdus, Md Sayem Ul Haque, Radha Das, & Sadia Sultana. (2024). COMPARATIVE ANALYSIS OF SENTIMENT ANALYSIS MODELS ON BANKING INVESTMENT IMPACT BY MACHINE LEARNING ALGORITHM. International Journal of Computer Science & Information System, 9(11), 5–16. https://doi.org/10.55640/ijcsis/Volume09Issue11-02

4. Das, A. C., Mozumder, M. S. A., Hasan, M. A., Bhuiyan, M., Islam, M. R., Hossain, M. N., ... & Alam, M. I. (2024). MACHINE LEARNING APPROACHES FOR DEMAND FORECASTING: THE IMPACT OF CUSTOMER SATISFACTION ON PREDICTION ACCURACY. The American Journal of Engineering and Technology, 6(10), 42-53.

5. Akter, S., Mahmud, F., Rahman, T., Ahmmed, M. J., Uddin, M. K., Alam, M. I., ... & Jui, A. H. (2024). A COMPREHENSIVE STUDY OF MACHINE LEARNING APPROACHES FOR CUSTOMER SENTIMENT ANALYSIS IN BANKING SECTOR. The American Journal of Engineering and Technology, 6(10), 100-111.

6. Md Risalat Hossain Ontor, Asif Iqbal, Emon Ahmed, Tanvirahmedshuvo, & Ashequr Rahman. (2024). LEVERAGING DIGITAL TRANSFORMATION AND SOCIAL MEDIA ANALYTICS FOR OPTIMIZING US FASHION BRANDS' PERFORMANCE: A MACHINE LEARNING APPROACH. International Journal of Computer Science & Information System, 9(11), 45–56. https://doi.org/10.55640/ijcsis/Volume09Issue11-05

7. Rahman, A., Iqbal, A., Ahmed, E., & Ontor, M. R. H. (2024). PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT. International journal of business and management sciences, 4(12), 18-32.

8. Md Jamil Ahmmed, Md Mohibur Rahman, Ashim Chandra Das, Pritom Das, Tamanna Pervin, Sadia Afrin, Sanjida Akter Tisha, Md Mehedi Hassan, & Nabila Rahman. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. International Journal of Computer Science & Information System, 9(11), 31–44. https://doi.org/10.55640/ijcsis/Volume09Issue11-04

9. Arif, M., Ahmed, M. P., Al Mamun, A., Uddin, M. K., Mahmud, F., Rahman, T., ... & Helal, M. (2024). DYNAMIC PRICING IN FINANCIAL TECHNOLOGY: EVALUATING MACHINE LEARNING SOLUTIONS FOR MARKET ADAPTABILITY. International Interdisciplinary Business Economics Advancement Journal, 5(10), 13-27.

10. Uddin, M. K., Akter, S., Das, P., Anjum, N., Akter, S., Alam, M., ... & Pervin, T. (2024). MACHINE LEARNING-BASED EARLY DETECTION OF KIDNEY DISEASE: A COMPARATIVE STUDY OF PREDICTION MODELS AND PERFORMANCE EVALUATION. International Journal of Medical Science and Public HealthResearch, 5(12),58-75.

11. Buhlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model selection. Statistical Science, 22(4), 477-504. https://doi.org/10.1214/07-STS241

12. Chen, Y., Zhang, T., & Xu, H. (2018). Forecasting stock returns with machine learning. Journal of Financial Data Science, 1(1), 1-26. https://doi.org/10.3905/jfds.2018.1.1.1

13. Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. Journal of Economic Perspectives, 15(4), 157-168. https://doi.org/10.1257/jep.15.4.157

14. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669. https://doi.org/10.1016/j.ejor.2018.03.024

15. He, K., Zhang, X., Ren, S., & Sun, J. (2017). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

16. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

17. Jiang, Z., Xu, Z., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. arXiv

preprint arXiv:1706.10059.

18. Li, Y., Li, Y., & Zhang, C. (2020). Transformer-based financial market prediction models: A review. Financial Innovation, 6(1), 1-14. https://doi.org/10.1186/s40854-020-00191-9

19. Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

20. Zhang, Z., Zhang, H., & Li, Q. (2019). Stock market forecasting with gradient boosting machines. Journal of Computational Finance, 23(2), 89-114. https://doi.org/10.21314/JCF.2019.197

21. Shak, M. S., Uddin, A., Rahman, M. H., Anjum, N., Al Bony, M. N. V., Alam, M., ... & Pervin, T. (2024). INNOVATIVE MACHINE LEARNING APPROACHES TO FOSTER FINANCIAL INCLUSION IN MICROFINANCE. International Interdisciplinary Business Economics Advancement Journal, 5(11), 6-20.

22. Naznin, R., Sarkar, M. A. I., Asaduzzaman, M., Akter, S., Mou, S. N., Miah, M. R., ... & Sajal, A. (2024). ENHANCING SMALL BUSINESS MANAGEMENT THROUGH MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS FOR CUSTOMER RETENTION, FINANCIAL FORECASTING, AND INVENTORY OPTIMIZATION. International Interdisciplinary Business Economics Advancement Journal, 5(11), 21-32.

23. Bhattacharjee, B., Mou, S. N., Hossain, M. S., Rahman, M. K., Hassan, M. M., Rahman, N., ... & Haque, M. S. U. (2024). MACHINE LEARNING FOR COST ESTIMATION AND FORECASTING IN BANKING: A COMPARATIVE ANALYSIS OF ALGORITHMS. Frontline Marketing, Management and Economics Journal, 4(12), 66-83.

24. Rahman, A., Iqbal, A., Ahmed, E., & Ontor, M. R. H. (2024). PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT. Frontline Marketing, Management and Economics Journal, 4(12), 84-106.

25. Al Mamun, A., Hossain, M. S., Rishad, S. S. I., Rahman, M. M., Shakil, F., Choudhury, M. Z. M. E., ... & Sultana, S. (2024). MACHINE LEARNING FOR STOCK MARKET SECURITY MEASUREMENT: A COMPARATIVE ANALYSIS OF SUPERVISED, UNSUPERVISED, AND DEEP LEARNING MODELS. The American Journal of Engineering and Technology, 6(11), 63-76.

26. Das, A. C., Rishad, S. S. I., Akter, P., Tisha, S. A., Afrin, S., Shakil, F., ... & Rahman, M. M. (2024). ENHANCING BLOCKCHAIN SECURITY WITH MACHINE LEARNING: A COMPREHENSIVE STUDY OF ALGORITHMS AND APPLICATIONS. The American Journal of Engineering and Technology, 6(12), 150-162.

27. Miah, J., Khan, R. H., Ahmed, S., & Mahmud, M. I. (2023, June). A comparative study of detecting covid 19 by using chest X-ray images–A deep learning approach. In 2023 IEEE World AI IoT Congress (AIIoT) (pp. 0311-0316). IEEE.

28. Khan, R. H., Miah, J., Nipun, S. A. A., & Islam, M. (2023, March). A Comparative Study of Machine Learning classifiers to analyze the Precision of Myocardial Infarction prediction. In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0949-0954). IEEE.

29. Kayyum, S., Miah, J., Shadaab, A., Islam, M. M., Islam, M., Nipun, S. A. A., ... & Al Faisal, F. (2020, January). Data analysis on myocardial infarction with the help of machine learning algorithms considering distinctive or non-distinctive features. In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-7). IEEE.

30. Islam, M. M., Nipun, S. A. A., Islam, M., Rahat, M. A. R., Miah, J., Kayyum, S., ... & Al Faisal, F. (2020). An empirical study to predict myocardial infarction using k-means and hierarchical clustering. In Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2 (pp. 120-130). Springer Singapore.

31. Miah, J., Ca, D. M., Sayed, M. A., Lipu, E. R., Mahmud, F., & Arafat, S. Y. (2023, November). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction. In 2023 15th International Conference on Innovations in Information Technology (IIT) (pp. 49-54). IEEE.

32. Khan, R. H., Miah, J., Rahat, M. A. R., Ahmed, A. H., Shahriyar, M. A., & Lipu, E. R. (2023, September). A Comparative Analysis of Machine Learning Approaches for Chronic Kidney Disease Detection. In 2023 8th

International Conference on Electrical, Electronics and Information Engineering (ICEEIE) (pp. 1-6). IEEE.

33. Miah, J., Cao, D. M., Sayed, M. A., Taluckder, M. S., Haque, M. S., & Mahmud, F. (2023). Advancing Brain Tumor Detection: A Thorough Investigation of CNNs, Clustering, and SoftMax Classification in the Analysis of MRI Images. arXiv preprint arXiv:2310.17720.

34. Rahman, M. M., Islam, A. M., Miah, J., Ahmad, S., & Mamun, M. (2023, June). sleepWell: Stress Level Prediction Through Sleep Data. Are You Stressed?. In 2023 IEEE World AI IoT Congress (AIIoT) (pp. 0229-0235). IEEE.

35. Rahman, M. M., Islam, A. M., Miah, J., Ahmad, S., & Hasan, M. M. (2023, June). Empirical Analysis with Component Decomposition Methods for Cervical Cancer Risk Assessment. In 2023 IEEE World AI IoT Congress (AIIoT) (pp. 0513-0519). IEEE.

36. Khan, R. H., Miah, J., Nipun, S. A. A., Islam, M., Amin, M. S., & Taluckder, M. S. (2023, September). Enhancing Lung Cancer Diagnosis with Machine Learning Methods and Systematic Review Synthesis. In 2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE) (pp. 1-5). IEEE.

37. Miah, J. (2024). HOW FAMILY DNA CAN CAUSE LUNG CANCER USING MACHINE LEARNING. International Journal of Medical Science and Public Health Research, 5(12), 8-14.

38. Miah, J., Khan, R. H., Linkon, A. A., Bhuiyan, M. S., Jewel, R. M., Ayon, E. H., ... & Tanvir Islam, M. (2024). Developing a Deep Learning Methodology to Anticipate the Onset of Diabetic Retinopathy at an Early Stage. In Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency: Volume 1 (pp. 77-91). Cham: Springer Nature Switzerland.

39. Rahman, M. M., Akhi, S. S., Hossain, S., Ayub, M. I., Siddique, M. T., Nath, A., ... & Hassan, M. M. (2024). EVALUATING MACHINE LEARNING MODELS FOR OPTIMAL CUSTOMER SEGMENTATION IN BANKING: A COMPARATIVE STUDY. The American Journal of Engineering and Technology, 6(12), 68-83.

40. Das, P., Pervin, T., Bhattacharjee, B., Karim, M. R., Sultana, N., Khan, M. S., ... & Kamruzzaman, F. N. U. (2024). OPTIMIZING REAL-TIME DYNAMIC PRICING STRATEGIES IN RETAIL AND E-COMMERCE USING MACHINE LEARNING MODELS. The American Journal of Engineering and Technology, 6(12), 163-177.

41. Hossain, M. N., Hossain, S., Nath, A., Nath, P. C., Ayub, M. I., Hassan, M. M., ... & Rasel, M. (2024). ENHANCED BANKING FRAUD DETECTION: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS. American Research Index Library, 23-35.

42. Shak, M. S., Mozumder, M. S. A., Hasan, M. A., Das, A. C., Miah, M. R., Akter, S., & Hossain, M. N. (2024). OPTIMIZING RETAIL DEMAND FORECASTING: A PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS INCLUDING LSTM AND GRADIENT BOOSTING. The American Journal of Engineering and Technology, 6(09), 67-80.

43. Das, A. C., Mozumder, M. S. A., Hasan, M. A., Bhuiyan, M., Islam, M. R., Hossain, M. N., ... & Alam, M. I. (2024). MACHINE LEARNING APPROACHES FOR DEMAND FORECASTING: THE IMPACT OF CUSTOMER SATISFACTION ON PREDICTION ACCURACY. The American Journal of Engineering and Technology, 6(10), 42-53.

44. Hossain, M. N., Anjum, N., Alam, M., Rahman, M. H., Taluckder, M. S., Al Bony, M. N. V., ... & Jui, A. H. (2024). PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR LUNG CANCER PREDICTION: A COMPARATIVE STUDY. International Journal of Medical Science and Public Health Research, 5(11), 41-55.

45. Ahmmed, M. J., Rahman, M. M., Das, A. C., Das, P., Pervin, T., Afrin, S., ... & Rahman, N. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. American Research Index Library, 31-44.

46. Al Bony, M. N. V., Das, P., Pervin, T., Shak, M. S., Akter, S., Anjum, N., ... & Rahman, M. K. (2024). COMPARATIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BUSINESS INTELLIGENCE: A STUDY ON CLASSIFICATION AND REGRESSION MODELS. Frontline Marketing, Management and Economics Journal, 4(11), 72-92.

47. Hasan, M., Pathan, M. K. M., & Kabir, M. F. (2024). Functionalized Mesoporous Silica Nanoparticles as Potential Drug Delivery Vehicle against Colorectal Cancer. Journal of Medical and Health Studies, 5(3), 56-62.

48. Hasan, M. (2023). SURFACE-ENGINEERED MINERAL PARTICLES FOR GATED DRUG DELIVERY, GENE TRANSFER AND SUNSCREEN FORMULATIONS.

49. Hasan, M., Kabir, M. F., & Pathan, M. K. M. (2024). PEGylation of Mesoporous Silica Nanoparticles for Drug Delivery Applications. Journal of Chemistry Studies, 3(2), 01-06.

50. Hasan, M., Evett, C. G., & Burton, J. (2024). Advances in Nanoparticle-Based Targeted Drug Delivery Systems for Colorectal Cancer Therapy: A Review. arXiv preprint arXiv:2409.05222.

51. Hasan, M., & Mahama, M. T. (2024). Uncovering the complex mechanisms behind nanomaterials-based plasmon-driven photocatalysis through the utilization of Surface-Enhanced Raman Spectroscopies. arXiv preprint arXiv:2408.13927.