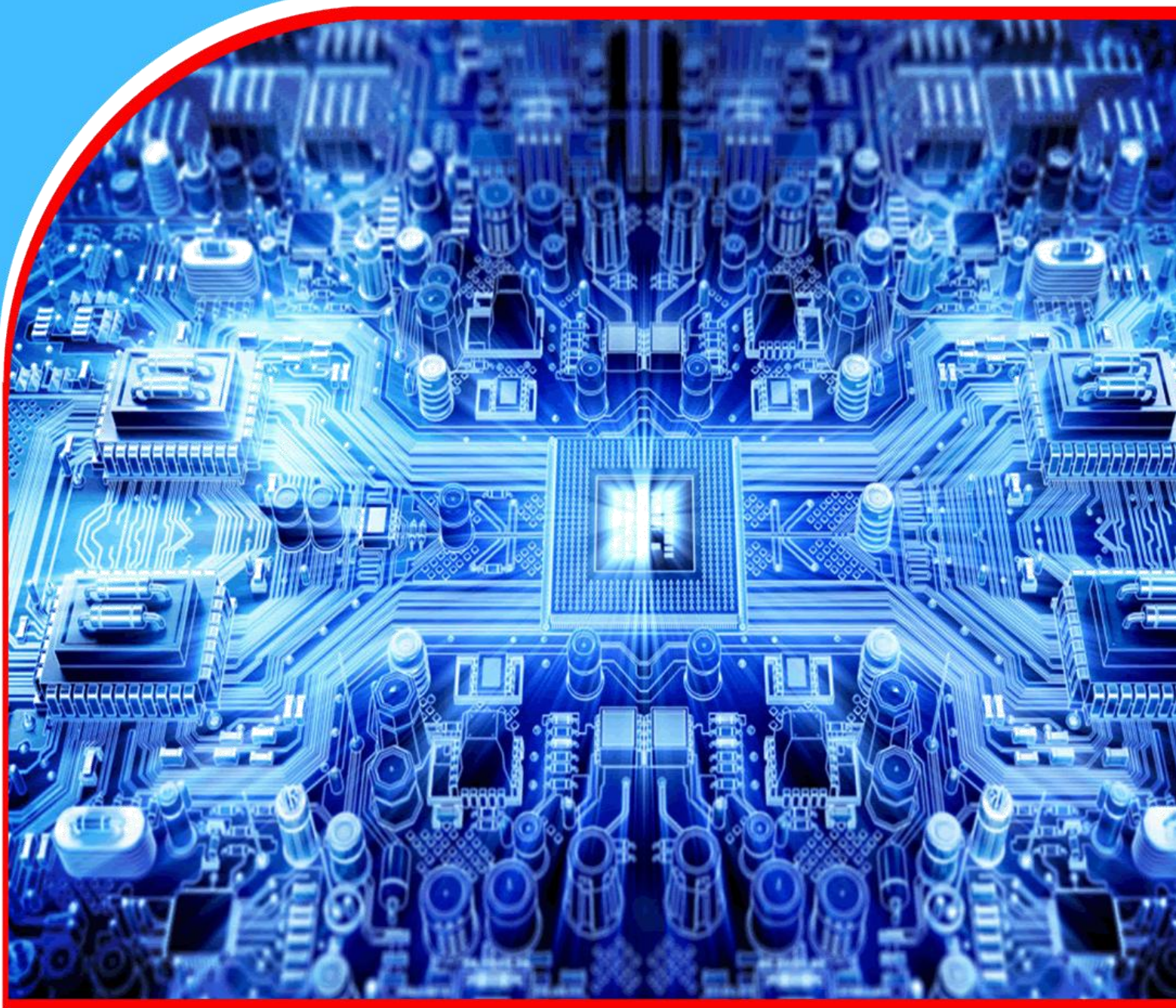
American Journal of
Computing and Engineering
(AJCE)



Effective Machine Learning Methods Used for Stock Price Forecasting: A Performance Evaluation

Khaja Moiezuddin Mohammed, Venkata Phaneendra Reddi, Pradeep Rao Vennamaneni, Sri Prudhvi Adusumilli, Rashmika Jalagam, Sridevi Allu

AJP

American Journal of Computing and Engineering
ISSN 2790-5586 (Online)
Vol.9, Issue 1, pp 1 - 18, 2026
www.ajpojournals.org

Effective Machine Learning Methods Used for Stock Price Forecasting: A Performance Evaluation

Khaja Moiezuddin Mohammed[1*], Venkata Phaneendra Reddi[2], Pradeep Rao Vennamaneni[3], Sri Prudhvi Adusumilli[4], Rashmika Jalagam[5], Sridevi Allu[6]
[1]Iqvia - Sr Data Engineer
[2]Atlantic Health systems - Oracle Integrations Developer
[3]JPMC - Senior Software Engineer
[4]Charter Communications, Sr Software Developer
[5]Matilda Cloud - Cloud Engineer/Python Developer
[6]American Airlines, AI/ML Engineer

Article history

Submitted 20.09.2026 Revised Version Received 17.10.2026 Accepted 19.11.2026

Abstract

A crucial component of financial decision-making and investment planning is stock price forecasting, but it is a complicated problem because of the non-linear and unstable dynamics of the market. In this paper, the researcher outlines a functioning stock market forecasting model which aims at enhancing the degree of forecasting precision founded on the past data of the S and P500 index. The approach involves the use of systematic data gathering and a thorough preprocessing procedure, such as the management of missing data, feature scaling with standardization and chronologically dividing the dataset into training (80%) and testing (20%) sets. Machine learning (ML) models called Random Forest (RF) and Linear Regression (LR) are used to analyze current market trends and forecast future prices. The key error measures (R2, MSE, RMSE and MAE) are used to evaluate the performance of each model, which guarantees an accurate evaluation of forecasting reliability. The results of the experiments prove that both models provide very accurate predictions, though LR has slightly better error performance, and RF has high levels of generalization. The comparative analysis also demonstrates that proposed models are superior to the current models, such as ANN and SVM. Overall, the work offers an informative, evidence-based prediction model, which is resistant, stable, and can be used in supporting real-time financial decisions and investment policies.

Keywords: *Stock Price Forecasting, S&P 500 Index, Machine Learning, Financial Market Prediction, Linear Regression, Random Forest, Stock Market Analysis.*

JEL Codes: *C52, C53, C58, G17*

I. INTRODUCTION

The ambition to outperform other market players is not a recent development in the history of financial markets, and the competence to predict time series adequately is indeed an issue that the market participants are interested in continuously. In the dynamic markets, financial prediction and decision making are typically based on stock price forecasting, which plays a significant role in investment choices, risk management as well as supporting strategic plans [1]. In order to be situated at the core, forecasting depends on the past market action, price deformities, trade volume and in general, economic indicators, which cumulatively affect the future stock movements. However, stock markets are nonlinear and volatile in nature and open to a wide range of extraneous factors which involve geopolitics, investor mood and unforeseen shocks in the market [2]. All these make it hard to make accurate forecasts as well as applying traditional statistical or rule-based methods to attain this is all but impossible. In order to tackle these challenges, the modern forecasting systems increasingly revolve around the application of structured information processing and feature design, identification of trend patterns and time-series knowledge [3][4]. Forecasting systems are used to explain the underlying market processes through a keen examination of the historical trends of the market, market indicators and time dependencies. All these elements are linked to form a solid analytical foundation under which the quality of data, the reduction of noise, and the identification of patterns is significant in enhancing predictive capability [5]. The dynamic nature of markets also demands scaling and flexible strategies which can have the ability to obtain new information and respond to the rapid changes in the markets.

In this regard, Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are the most recent innovative technologies that can radically change the extent of stock price forecasting [6]. Smart systems that can unravel complex financial situations are brought by AI. ML assists in the fact that it enables the development of the prediction techniques that rely on the data to learn the relations relying on the historical data, without any human intervention [7]. This field is also developed by DL as it is able to find rich hierarchical structures and make long term discoveries about sequential stock data [8]. AI, ML, and DL may be integrated into an effective and connected system which makes the stock prices even more accurate, flexible and capable of forecasting them on a considerably higher level than the conventional paradigms in forecasting. The key contributions are:

- A long-term stock price prediction model is developed by assistance of strict preprocessing, normalization of features and an optimum train-test strategy to ensure trusted performance of the models.
- LR and RF: Implementation and comparison analysis provides an ideal overview of how linear and nonlinear learning processes may be applied to understand the market behavior.
- Prediction accuracy and model stability can be gauged on a good quantitative basis using an elaborate performance evaluation in terms of R 2, MSE, RMSE, and MAE.
- The work has given a clear computationally efficient and generalizable forecasting strategy that can be applied in practice in finance and investment industries.

A. Justification and Novelty of Paper

This research is justified by the fact that there is an increased demand in reliable, data-driven forecasting techniques that can overcome the volatility which is inherent in the financial markets and help make informed investment choices. Whereas the traditional frameworks tend to fail in dealing with non-linear market dynamics, the proposed framework presents a systematic approach, involving rigorous preprocessing, the standardization of feature scaling, and the

comparative application of two ML algorithms with high performance, RF and LR are used to simulate both linear and non-linear market dynamics. The originality of this work lies in its methodical combination of the precision of feature-engineering with the quality measurement of evaluation, which presupposes the possibility of forecasting the effect on S&P 500 stocks with great accuracy. This model provides high performance and computational simplicity, high transparency and strong generalization in contrast to the previous studies which have high complex architectures with low interpretability. This makes the framework efficient, accessible and ideal to real-time financial analytics and decision-support applications.

B. Organization of the Paper

The structure of this paper is as follows: Section II discuss Related Work reviews previous studies on stock forecasting using ML. Section III Methodology describes the whole approach that was used for building and applying forecasting models. Section IV exhibits the Results and Discussion which presents and analyzes the outcomes of the implemented models. Section V explains the Conclusion and further Scope highlights important discoveries and makes recommendations for further research.

II. Literature Review

This section examines and highlights estimate techniques and predictive analytics, with a focus on stock price predictions based on ML. Methodologically, the field has evolved from traditional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) [22], [23], which assume linearity and stationarity in the underlying series, to ML models such as SVM and ANN that learn non-linear relationships directly from data, and more recently to ensemble techniques such as Random Forest [24] that combine many learners to improve stability and generalization. Despite this evolution, existing studies differ widely in datasets, features, data-splitting schemes and evaluation metrics, which makes their results difficult to compare or reproduce. The studies reviewed below are discussed in this context. Among the works that have been assessed are:

Kalra and Prasad (2019) propose an algorithm that forecasts changes in the Indian stock market every day by utilizing news articles and historical data. The classifier is used to categorize the news content as either positive or negative. Bayes Naïve. For forecast purposes, the number of news stories with positive and negative sentiment for each day, the difference in the close prices of neighboring days, and historical data are employed. A range of 65.30 to 91.2% accuracy is attained using different machine learning algorithms [9].

Moni (2019) aimed to forecast the yearly S&P 500 stock market index in order to produce an indication of political stability on a worldwide scale. Using machine learning and genetic programming, an algorithm was developed with a template that took into account the Dow Jones Industrial Average, Crude oil prices, gold prices, the number of American military deaths, and the S&P 500 stock index, and inflation rates in the United States. This algorithm produced a very accurate forecast, within 14% [10].

Pun and Shahi (2018) For sector-wise analysis, Ten different investment sectors are used to further categorise stock data. The stock price is predicted for the following day using support vector regression (SVR) and artificial neural networks (ANN). To evaluate two learning models, the mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R2) are employed. The result shows that, except for mutual funds, development banks, and the financial sector, SVR with min max normalization performs better than ANN across all industries [11].

Tantisripreecha and Soonthomphisaj (2018). Linear Discriminant Analysis (LDA). The feature set that is utilized includes the beginning, ending, peak, and minimum prices. According to experiments conducted on four popular NASDAQ stocks; Apple, Facebook, Google, and Amazon the method yields the best stock prediction outcomes. In both batch and online learning methodologies, compare LDA-online with ANN, KNN, and Decision Tree. It was found that LDA-Online produced the best outcomes. The stocks of Google, Amazon, Apple, and Facebook had the biggest returns, at 97.81%, 97.64%, 95.58%, and 95.18%, respectively [12].

Batra and Daudpota (2018) assessed the attitude of tweets on Apple products that were extracted from StockTwits, a social media platform, between 2010 and 2017. It has used tweets as well as market index data for the same time period that was extracted from Yahoo Finance. An SVM model is then constructed using the sentiment score and market data to forecast the movement of stocks the next day. The results show that market data and public opinion may be positively correlated, and the recommended work's stock prediction accuracy is 76.65% [13].

Maini and Govinda (2017) suggests utilising ML algorithms such as the RF model and SVM to forecast stock market movements. An incredibly effective ensemble learning technique for regression and classification is the RF model. One type of ML model for classification is the SVM. But categorisation is the primary use for this technique [14].

Li and Liao (2017) assesses Recurrent Neural Network (RNN), Multilayer Perceptron (MLP), Forecasting models include Support Vector Machine (SVM), Decision Tree, Naive Bayes, and Long Short-Term Memory (LSTM). To assess the models' performance in diverse circumstances, trainsets of 12 distinct record numbers are chosen, and nine feature combinations are chosen based on 23 technical indicators that are often utilized in stock market analysis. Data spanning 8 years, from 2008 to 2015, are used to study the Chinese stock market listed corporation (000592). According to experimental findings, the DL models MLP, RNN, and LSTM have greater accuracy indices than other models. Decision Tree performs better than MLP by 20.75%, better than others according to the f-measure, and better than Naive Bayes by 40.02% [15].

Suksiri, Chiewchanwattana and Sunat (2016) suggests a stock prediction model's preprocessing method is a single spectrum analysis-based kernel-based extreme learning machine (SSA-KELM). Five different models are compared against SSA-KELM's prediction ability. Performance is assessed using three stock price data points. The experimental data shows that all SSA-based models can perform better than non-SSA models. Additionally, compared to previous SSA-based models, the results suggest that the quickest training time and the highest accuracy could be achieved with SSA-KELM [16].

Taken together, these studies show a clear progression from statistical models to ML and ensemble methods; however, most of them rely on different datasets, evaluate on non-chronological or unspecified splits, and report inconsistent metrics, and very few provide a reproducible baseline evaluation on S&P 500 data. This gap motivates the present study, which benchmarks LR and RF on S&P 500 data under a chronological train-test protocol with standard error metrics.

Table I provides a comparative overview of recent research studies on Stock Price Forecasting using various ML models. It summarizes each study’s methodology, dataset, performance, limitations, and suggested future work

Table 1: Summary on Machine Learning Approaches for Stock Price Forecasting

Reference	Methodology	Dataset	Performance	Limitations	Future Work
Kalra & Prasad (2019)	Sentiment-driven stock movement prediction using Naïve Bayes classifier to classify news as positive/negative; combined sentiment counts with historical stock data.	Indian stock market data + company news articles	Accuracy between 65.30%-91.2% across ML models	Relies heavily on sentiment classification quality; limited feature diversity; no deep learning models applied.	Incorporate deep learning-based sentiment analysis (BiLSTM, transformers) and more market indicators for improved accuracy.
Moni (2019)	Genetic Programming (GP) based model using multiple macroeconomic and geopolitical features to predict annual S&P 500 index.	Historical S&P 500, gold prices, oil prices, casualties in wars, inflation, DJIA	Prediction accuracy within 14% error margin	Low temporal granularity (annual data only); GP can overfit; lacks real-time prediction capability.	Extend to monthly/daily forecasting; integrate advanced evolutionary algorithms and deep learning.
Pun & Shahi (2018)	Stock prediction using SVR and ANN with Min-Max normalization; performance measured via MSE, MAE, RMSE, R^2.	Ten sector-wise stock data	SVR outperforms ANN across the majority of industries, with the exception of development banks, finance, and mutual funds	Limited to next-day prediction; feature set not extensive; ANN architecture not optimized.	Introduce hybrid DL models (LSTM-CNN); include technical indicators; improve sector-wise tuning.
Tantisripreecha &	Linear Discriminant Analysis	NASDAQ: Apple,	Google, Amazon, Apple, and	Only OHLC features used; lacks	Add hybrid OHLC + technical +

Reference	Methodology	Dataset	Performance	Limitations	Future Work
Soonthomphisaj (2018)	(LDA-Online) using OHLC features; compared with ANN, KNN, DT in batch and online mode.	Facebook, Google, Amazon	Facebook had accuracy rates of 97.81%, 97.64%, 95.58%, and 95.18%, respectively.	sentiment/technical indicators; model may not generalize to volatile markets.	sentiment features; test on more global markets.
Batra & Daudpota (2018)	Sentiment analysis of StockTwits tweets using SVM; tweets labeled bullish/bearish and combined with market index data for next-day prediction.	Apple tweets (2010–2017) + Yahoo Finance index data	Stock movement prediction accuracy 76.65%	Single-company dataset; sentiment accuracy dependent on SVM; ignores advanced NLP models.	Use transformer-based embeddings (BERT/FinBERT); include multi-company and multi-source social media data.
Maini & Govinda (2017)	Prediction of market trends using Random Forest and SVM models for classification.	Historical market datasets (not explicitly detailed)	Good classification performance reported qualitatively	Lacks detailed performance metrics; no comparison with deep learning.	Extend study with DL (LSTM/GRU) and include time-series indicators and sentiment factors.
Li & Liao (2017)	Evaluated 23 technical indicators over 9 feature combinations and training with various record lengths to compare SVM, Naïve Bayes, Decision	Chinese stock market (000592), 8-year historical data (2008–2015)	Deep learning models perform better than others; in the F-measure, MLP beats Decision Tree by 20.75%, while Decision Tree beats Naïve	Computationally heavy; single stock; potential overfitting with deep models.	Apply to multiple stocks; incorporate news/sentiment data; use improved sequence models like Transformer.

Reference	Methodology	Dataset	Performance	Limitations	Future Work
	Tree, MLP, RNN, and LSTM.		Bayes by 40.02%		
Suksiri, Chiewchanwattana & Sunat (2016)	A comparison was made between five baseline models and the proposed single spectrum analysis preprocessing kernel extreme learning machine (SSA-KELM).	Three stock price datasets	SSA-KELM performs better than the other models in terms of training time and accuracy.	Limited features (mostly numerical time-series); SSA parameter tuning sensitive; no external indicators.	Add technical/sentiment indicators; explore hybrid SSA-deep learning models.

III. MATERIALS AND METHODS

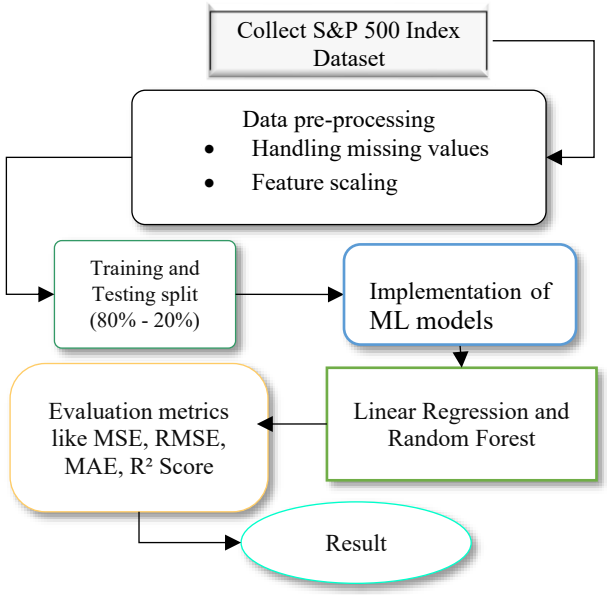


Figure 1: Flowchart for Stock Market Forecasting

A. Data Collection

The stock market prediction methodology, depicted in Figure. 1, starts with the gathering of past S&P 500 index data, which is the basic input for the forecasting pipeline. The data goes through a rigorous preprocessing stage during which missing values are tackled, and numerical attributes

are standardized through feature-scaling techniques. Subsets of the data are divided chronologically at an 80:20 ratio between training (earlier data) and testing (later data) following preprocessing to offer a dispassionate assessment of the model's functionality. The ML models, which are of the kind that learn market patterns and the temporal relationships underlying the market, are then used during the training phase. Metrics like MSE, RMSE, and MAE are used to evaluate the models' performance after training, as well as the coefficient of determination R2, which together gauge prediction accuracy and generalization capacity.

The S&P 500 stock market data offers five years of historical daily price information on all the companies that the NYSE collected and updated as of February 2018, and included in the S&P 500 index. Since every record has the required trade indicators, they provide an organized source of financial analytics, time-series forecasting studies, and ML-enhanced stock prediction research: Date, Open, Volume, Close, High, Low, and Stock Name.

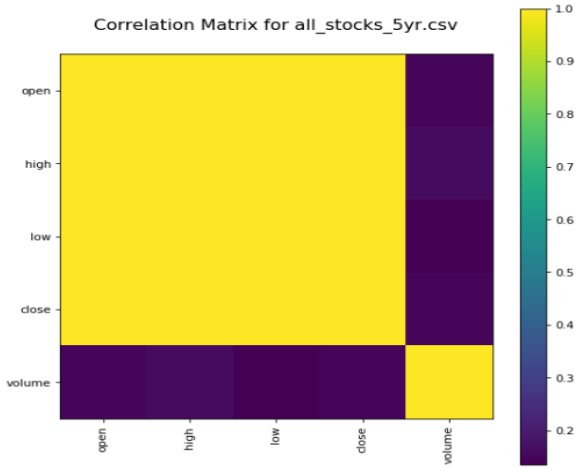


Figure 2: Correlation Matrix

In Figure 2 illustrate the correlation chart is used to visualize the linear correlation between the key elements of the opening, high, low, close, and volume of the stock market using five years' worth of S&P 500 trading data. The plot presents the positive correlation between the variables in price-related variables (Open, High, Low, Close) is extremely strong (near 1.0) meaning that the values change together closely on a day-to-day basis which is anticipated in data of financial time-series where the variables are dependent on each other i.e. price metrics. The overview of the correlation of the stock prices in this matrix is clear and at the same time illustrates the autonomy of the trading volume to the daily change in prices.

B. Data Preprocessing

Pre-processing is the process of converting raw data into a format that machine learning models can use. These involve Cleaning and converting data, Data sampling, managing missing values and scalability and feature engineering. These techniques enhance data quality, offer consistency, and help the model train more effectively. The key preprocessing actions that were used in this work are as follows:

1) Handling missing values

It is important to locate and make careful attention to missing values in a dataset. One approach to missing data that is used to impute data is data imputation which is using the data's context to estimate the missing information.

2) Feature Scaling Using Standardization

Features or standardization Scaling ensures that data is normalized within a range, making it a crucial step in data preparation. This is much faster in the calculation of algorithms [17]. StandardScaler method is applied in this work to feature scale. The transformation in data distribution is done with the help of StandardScaler where the mean is subtracted from each sample and the result is divided by the standard deviation. In this procedure, the data is scaled to a standard deviation of 1 and centered around 0. Equation (1) may be used to express the StandardScaler formula.

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

Where the original value is denoted by x and the scaled value by z. μ represents the characteristic's average value.σ is a representation of the feature values' standard deviation.

C. Data Splitting

The data was split chronologically into training (80%) and testing (20%) sets, with the earliest 80% of the observations used for training and the most recent 20% reserved for testing (training on the past and testing on the future). Chronological splitting, rather than random shuffling, prevents look-ahead bias and time-series data leakage. This stage made sure that previously unknown data was used for model evaluation.

D. Proposed with Machine Learning Models

After completing all preprocessing steps, the completed dataset was utilized to train several ML models. Among these were RF and LR.

1) Linear Regression

A dependent variable can be approximated using the linear regression (LR) approach, which reports the values or changes of other variables under study in a linear fashion. Assume that x_i and y_i are two variables in a continuous and random distribution. A numerical data set mapped by (x_i, y_i) for ii=1,2, ⋯, n, is assumed in this mode, where $x_i \in U^n$and $y_i \in U^n$. The linear relationship is a plausible type of relationship between the regressor X and the response variable Y [6]. In this way, a model may be shown as follows: $\Upsilon \in \omega \subseteq R^P$, which is constructed by χ and Γ; $\ddot{\Upsilon}_l$ of dimension n is a mapping in space $R^n \rightarrow R$, recorded by a vector of analysis connected to metrics involved; Equation (2) illustrates the basic linear regression in this method:

$$\ddot{\Upsilon}_l = \chi + \Gamma X_l \qquad (2)$$

where χ represents the value at which Xl=0, also known as the intersection, and Γ represents the change in $\ddot{\Upsilon}_l$, also known as the line's slope.

2) Random Forest

A collection of unpruned regression or classification trees made by choosing features at random from bootstrap samples of the training data is called a random forest [18]. Exploratory variables do not need to have a functional form or be multicollinear in order to use classifier trees. Because they employ non-parametric methods, no distributional assumptions are required. Below is given Equation (3), which represents Random Forest mathematically.

$$\widehat{f_{bag}} = \frac{1}{B}\sum_{b=1}^{B} \widehat{f^b}(x) \qquad (3)$$

In the given formula, $\widehat{f_{bag}}$ represents the final prediction made by the Random Forest model for a given input x. The variable B denotes the total number of decision trees (or base learners) used

in the ensemble. Each individual tree produces a prediction denoted by $\widehat{f^b}(x)$, where b ranges from 1 to B.

E. Performance Metrics

Performance metrics such as R^2, MSE, RMSE, and MAE is essential for evaluating model accuracy in stock price forecasting. R^2 indicates variance explanation, MSE and RMSE assess squared prediction errors, and MAE measures average absolute errors [19]. Together, they offer a comprehensive evaluation of a model's reliability and prediction accuracy.

R^2: A crucial statistic used to quantify the percentage of the discrepancy in the coefficient of determination is the dependent variable that the independent factors can explain. Equation (4) represents the formulation of R^2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (4)$$

MSE: The average of actual and expected data is obtained using mean square error, which also removes values that are less than zero. Equation (5) provides the MSE formulation.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_{u,i} - r_{u,i})^2 \qquad (5)$$

where $r_{u,i}$ is the related actual rating and $P_{u,i}$ is the user u's anticipated rating of service i.

RMSE: The RMSE, which is determined by Equation (6), is the square root of the mean squared difference between the actual output and the expected output. For extremely non-linear errors, it is employed. A useful indicator of prediction accuracy, the average amount of errors in the simulated data is displayed by the RMSE.

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{(SWL_{FOR,i} - SWL_{OBS,i})}{N}} \qquad (6)$$

Where $SWL_{FOR,i}$ is the forecast output, $SWL_{OBS,i}$ is the actual output, N is the total number of data points, and I is a single data input from the data points.

MAE: The expected and actual values the average absolute deviation is measured as MAE. Its sensitivity to outliers is lower than that of RMSE. Smaller MAEs indicate less mistakes in predictions. MAE formulation is given in Equation (7).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{1} \qquad (7)$$

where yi stands for the observed SM, $\bar{y}$ for the observed SM mean, and $\hat{y}_i$ for the model-predicted SM.

IV. FINDINGS

This study analyzes the findings from proposed models applied to Forecast Stock Prices. Experiments were conducted using Python 3.7 with Random Forest, skLearn 1.6.1, and Linear Regression on an Intel Core i5 CPU, Intel(R), and 8 GB of RAM running 64-bit Windows 10. The results of the LR and RF to forecast stock prices based on four evaluation metrics are shown in Table 2, R2, MSE, RMSE, and MAE. Both models are very predictive, and the R2 is very high at 99.77 percent and 99.67 percent for Linear regression and random forest respectively, which is a strong fit in either of the two models. But the Linear Regression performs better by all error measures as it has lower MSE, RMSE and MAE which indicates that it makes more accurate and consistent predictions on such data.

Table 2: Performance of Linear Regression and Random Forest Models on S&P 500 stock market data

Performance Metrics	Linear Regression	Random Forest
R^2	99.77	99.69
MSE	370.05	533.58
RMSE	19.23	23.59
MAE	13.79	16.98

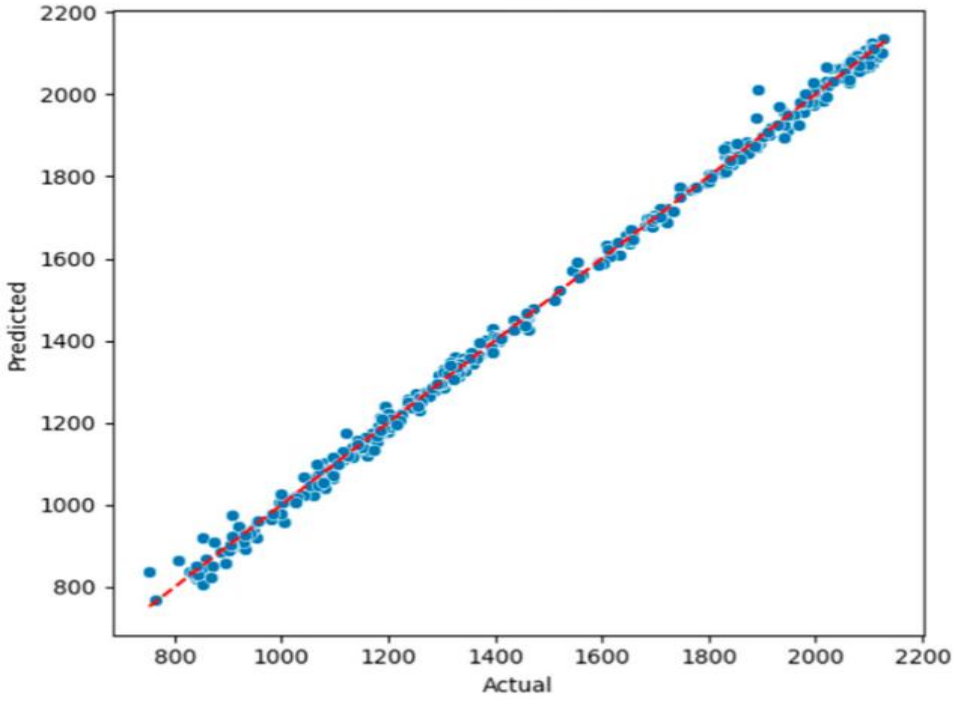


Figure 3: Actual vs. Predicted Values for Linear Regression Model

Figure 3 shows a scatter plot of the model values against the real stock values. Because the predicted values closely follow the actual values over the test period, the blue values lie close to the red dashed diagonal line, which indicates that the LR model is quite accurate in tracking short-term trends over this specific timeframe. The tight placement of the dots around the reference line demonstrates that, the errors of prediction are also very minor and indicate that the model also manages to learn and reproduce the underlying trend in the stock data.

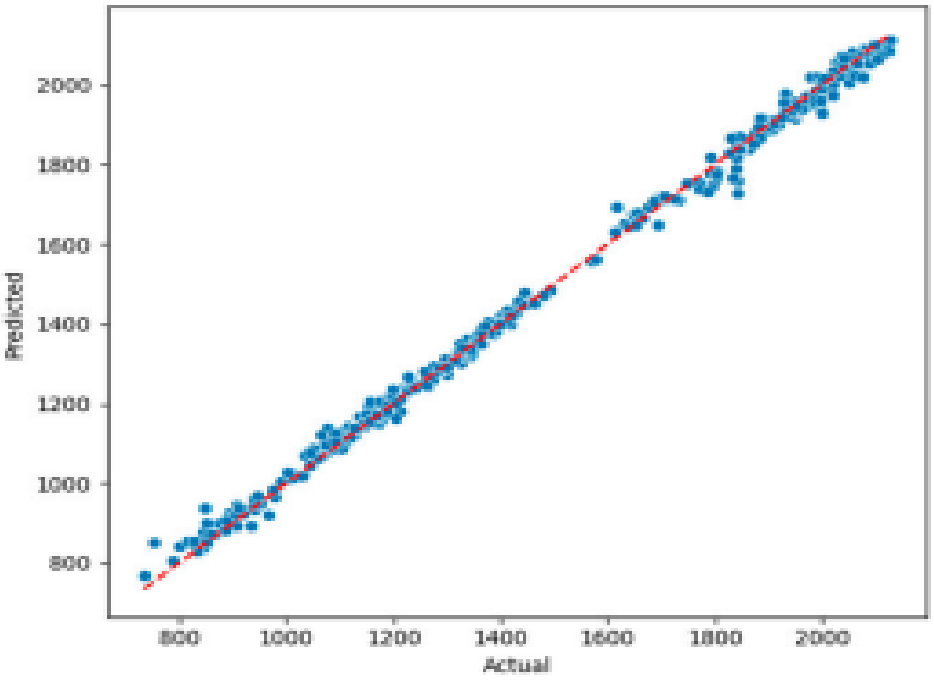


Figure 4: Actual vs. Predicted Values from Random Forest Model

The Figure 4 is a scatter plot that compares the actual stock values and the ones that the Random Forest model predicts. The blue points lie directly parallel to the red dashed diagonal line and this is to mean that the underlying trend is effectively captivated by the model and results in highly accurate predictions. Even though small deviation points can be observed, general high concentration of points proves that the model is highly working and can be used to make predictions on the movements of stock prices with high levels of generalization within the entire dataset and hence random forest is a good decision in terms of prediction of stock price movement.

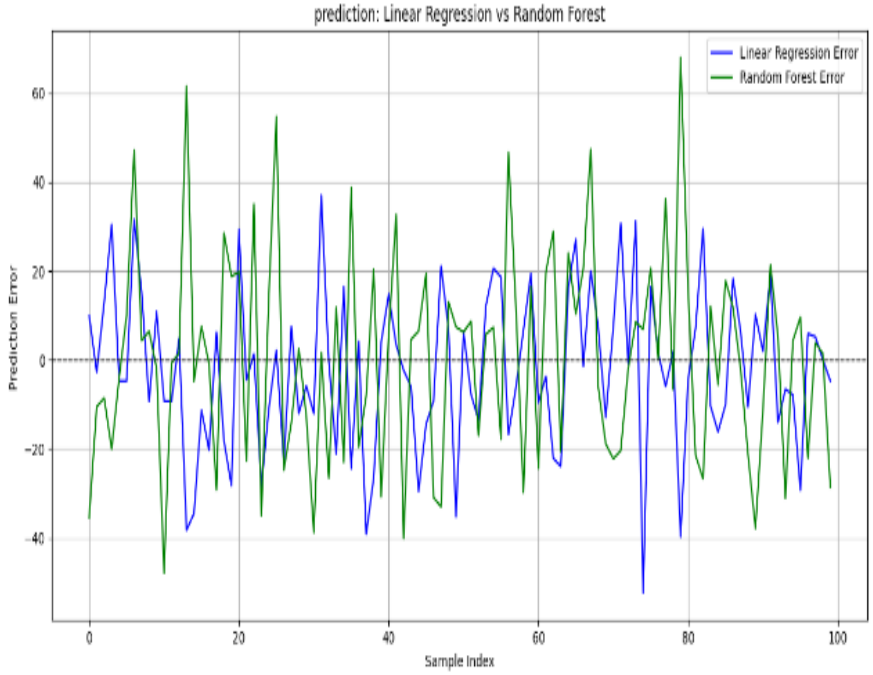


Figure 5: Prediction Error: Linear Regression vs Random Forest

In Figure 5 shows the number of samples used in this figure were 100 points, and the prediction errors between A comparison has been made between the RF and LR models. The blue line is an illustration of the error variability of the Linear Regression and the green line depicts the error variability of the Random Forest model. Both models have oscillation about 0 which means that there are cases of overestimation and underestimation. Typically, the Random Forest model, however, exhibit more consistent and stable variations of errors than the Linear Regression model, indicating higher consistency and lower general error of prediction.

A. Comparative Analysis

Table 3 give a brief comparison of proposed models and current models that employ R^2 values to forecast the stock price. LR and RF exhibit the greatest predictive accuracy with a R^2 score of 99.77 and 99.67 respectively, which is evidence of a good model. Comparatively, as it can be seen the existing models, like ANN (86.6) and SVM (60.2) have a relatively lower performance, and this illustrates how the recommended methods have increased their accuracy in stock price prediction.

Table 3: Comparison of Suggested and Current Stock Price Forecasting Models

Model	R^2
Linear Regression	99.77
Random Forest	99.67
ANN[20]	86.6
SVM[21]	60.2

The stock price forecasting model suggested is built on the principles of complex and nonlinear and dynamic prediction of the historical market information using the latest data-driven learning methods, making the model highly predictive and stable. The model is capable of minimizing noise, enhancing the pattern recognition and generalization to changing market conditions by incorporating sophisticated feature processing and trend analysis as well as adaptive learning of the noise. It is robust and has the ability to capture concealed dependencies that other techniques cannot capture resulting in improved predictive ability and faster convergence along with more resilience to volatility and this is its key strength in real time decision making in finance and investment policies.

B. Limitations and Future Work

The suggested forecasting framework is not a framework without limitations as far as it has performed well regarding predictions. The study relies purely on price related features in the previous years, and they may not always indicate all the market externalities such as the economic news, political relations, or investor activity that largely influences the flow of the stock. Further, the data sets are limited to a given time span of five years and therefore might not be as flexible to long term structural changes in the market behavior. The combination of multimodal information, including news sentiment, macroeconomic factors, and dynamic financial data may be discussed as the further research that can be undertaken to make the model better and mindful of the environment. Learning complex time-dependent relations can be also improved by applying deep learning structures, hybrid ensemble methods, and attention mechanisms. Moreover, in order to achieve greater generalizability and develop stronger stock prediction systems, there would be a requirement to evaluate the model's performance using bigger, more varied data sets and test the models over market regimes that vary, including turbulent, bearish, and bullish.

The superior error performance of LR should not, however, be interpreted as proof that the underlying price-generating process is linear. S&P 500 price levels are non-stationary, and

regressing non-stationary series on one another can produce a spuriously high R^2 and low error values without any genuine structural relationship [25]. Accordingly, although linear models perform well in short-term trend following over the specific timeframe analyzed here, they are likely to fail during regime shifts and high-volatility market crashes, when the relationships among variables change abruptly. Further validation using stationarity tests (e.g., the Augmented Dickey–Fuller test), differenced or return-based targets, and out-of-sample evaluation across multiple market regimes is therefore required before such models are relied upon in practice

V. Conclusion and Future Scope

Forecasting stock prices is a volatile and intricate process especially when it comes to indexes such as S and P 500 hence it is an important undertaking by investors and financial experts who want to know the direction that the market is heading. In the study, an ML-based, systematic, and data-driven stock price prediction framework of the S&P 500 was developed. To ultimately guarantee fairness in the model testing, it made use of feature scaling, chronological 80:20 train to test split, and thorough data pre-processing. The market's underlying trend was represented using Linear Regression and Random Forest, while R, MSE, RMSE, and MAE were utilized to assess the outcomes. The results of the studies shown that, with the aid of the two models, the prediction accuracy is very comparable since the Linear Regression produced a little better error performance, Nevertheless, the Random Forest is stable and has a strong ability for generalization. It was also found that the comparative analysis proved that these proposed models were superior to the existing methods such as ANN or SVM. Overall, the findings indicate that the specified forecasting model can be considered reliable, efficient, and capable of facilitating the improved financial planning, risk estimation, and investment decision-making in a real market environment.

REFERENCES

[1] L. Ryll and S. Seidens, "Evaluating the Performance of Machine Learning Algorithms in Financial Market Forecasting : A Comprehensive Survey," 2019, doi: 10.48550/arXiv.1906.07786.

[2] S. Sarode, H. G. Tolani, P. Kak, and C. S. Lifna, "Stock Price Prediction Using Machine Learning Techniques," in *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, 2019, pp. 177–181. doi: 10.1109/ISS1.2019.8907958.

[3] D. Selvamuthu, V. Kumar, and A. Mishra, "Indian stock market prediction using artificial neural networks on tick data," *Financ. Innov.*, vol. 5, no. 1, 2019, doi: 10.1186/s40854-019-0131-7.

[4] C.-F. Tsai and W. P, "Stock Price Forecasting by Hybrid Machine Learning Techniques," *Proc. Int. MultiConference Eng. Comput. Sci.*, vol. 1, 2009.

[5] K. C. Rasekhschaffe and R. C. Jones, "Machine Learning for Stock Selection," *Financ. Anal. J.*, vol. 75, no. 3, pp. 70–88, 2019, doi: 10.1080/0015198X.2019.1596678.

[6] I. Parmar *et al.*, "Stock Market Prediction Using Machine Learning," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, IEEE, Dec. 2018, pp. 574–576. doi: 10.1109/ICSCCC.2018.8703332.

[7] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms," *Intell. Syst. Accounting, Financ. Manag.*, vol. 26, no. 4, pp. 164–174, Oct. 2019, doi: 10.1002/isaf.1459.

[8] A. Ntakaris, M. Magris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *J. Forecast.*, vol. 37, no. 8, pp. 852–866, 2018, doi: 10.1002/for.2543.

[9] S. Kalra and J. S. Prasad, "Efficacy of News Sentiment for Stock Market Prediction," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 491–496. doi: 10.1109/COMITCon.2019.8862265.

[10] V. Moni, "Machine Learning to Predict Annual Stock Market Index - a Genetic Programming Approach," in *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, IEEE, Apr. 2019, pp. 1–4. doi: 10.1109/ICIICT1.2019.8741439.

[11] T. B. Pun and T. B. Shahi, "Nepal Stock Exchange Prediction Using Support Vector Regression and Neural Networks," in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, IEEE, Feb. 2018, pp. 1–6. doi: 10.1109/ICAECC.2018.8479456.

[12] T. Tantisripreecha and N. Soonthomphisaj, "Stock Market Movement Prediction using LDA-Online Learning Model," in *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, IEEE, Jun. 2018, pp. 135–139. doi: 10.1109/SNPD.2018.8441038.

[13] R. Batra and S. M. Daudpota, "Integrating StockTwits with sentiment analysis for better prediction of stock price movement," in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, IEEE, Mar. 2018, pp. 1–5. doi: 10.1109/ICOMET.2018.8346382.

[14] S. S. Maini and K. Govinda, "Stock market prediction using data mining techniques," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, Dec. 2017, pp. 654–661. doi: 10.1109/ISS1.2017.8389253.

[15] W. Li and J. Liao, "A comparative study on trend forecasting approach for stock price time series," in *2017 11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, IEEE, Oct. 2017, pp. 74–78. doi: 10.1109/ICASID.2017.8285747.

[16] P. Suksiri, S. Chiewchanwattana, and K. Sunat, "Application of singular spectrum analysis and kernel-based extreme learning machine for stock price prediction," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, Jul. 2016, pp. 1–6. doi: 10.1109/JCSSE.2016.7748873.

[17] M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, "Stock market prediction using machine learning techniques," in *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, IEEE, Aug. 2016, pp. 322–327. doi: 10.1109/ICCOINS.2016.7783235.

[18] S. Ray, "A Quick Review of Machine Learning Algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, IEEE, Feb. 2019, pp. 35–39. doi: 10.1109/COMITCon.2019.8862451.

[19] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018, doi: 10.1016/j.ejor.2017.11.054.

[20] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, 2015, doi: 10.1016/j.eswa.2014.07.040.

[21] N. Tripathy, "Stock Price Prediction Using Support Vector Machine Approach," in *Proceedings of The International Academic Conference on Management and Economics*, Acavent, Nov. 2019. doi: 10.33422/conferenceme.2019.11.641.

[22] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA, USA: Holden-Day, 1970.

[23] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, IEEE, Mar. 2014, pp. 106–112. doi: 10.1109/UKSim.2014.67.

[24] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[25] C. W. J. Granger and P. Newbold, "Spurious regressions in econometrics," *J. Econom.*, vol. 2, no. 2, pp. 111–120, 1974, doi: 10.1016/0304-4076(74)90034-7.

Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2022). Survey on Blockchain Integration with Sap S/4hana for Transparent and Secure Supply Chains. *International Journal of AI, BigData, Computational and Management Studies*, *3*(1), 161-170.

Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2022). A Review of Frameworks, and Impact on the Software Development Life Cycle with Artificial Intelligence in Software Engineering. *American International Journal of Computer Science and Technology*, *4*(1), 44-53.

Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2022). Review on Network Virtualization and SDN in Industrial IoT (IIoT). *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(1), 162-172.

Dasari, S., Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., & Pathani, A. K. (2022). Networking Paradigms for Large-Scale IoT Deployments: A Survey on Reliability, Latency, and Scalability. *International Journal of Emerging Trends in Computer Science and Information Technology*, *3*(1), 136-144.

Jannu, P. K., Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., & Pilli, J. (2022). Heat Dissipation Strategies for High-Performance Computing: A Review of Air and Liquid Cooling Efficiency. *International Journal of Emerging Trends in Computer Science and Information Technology*, *3*(1), 145-153.

Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2022). Towards Secure and Reliable IoT: A Review on Anomaly Detection and Predictive Analytics in Sensor Networks. *International Journal of AI, BigData, Computational and Management Studies*, *3*(1), 140-149.

Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., & Shyam,D. (2022). Understanding the Use of Agile Development Practices Alongside Quantum Computing in Modern Engineering Projects. *International Journal of Emerging Research in Engineering and Technology*, *3*(1), 136-143.

Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., Pilli, J., & Jannu, P. K. (2022). An Ensemble Machine Learning Approach for Reliable Fake News Detection on Social Media Platforms. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(1), 152-161.

Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2022). Comparative Performance of Machine Learning Techniques for Sentiment and Insight Analysis in Amazon Product Reviews. *American International Journal of Computer Science and Technology*, *4*(1), 54-64.

Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., & Keshireddy, S. R. (2022). Machine Learning-Driven Threat Analysis for Enhancing Security in Data Center Networking. *American International Journal of Computer Science and Technology*, *4*(1), 65-76.

Panchali, S. H., Mylavarapu, K. B., Pilli, J., Jannu, P. K., & Mohammad, J. A. (2022). Predictive Maintenance in Smart Manufacturing: A Review of Machine Learning and Digital Twin Approaches. *International Journal of AI, BigData, Computational and Management Studies*, *3*(1), 150-160.

Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., & Kavuluri, H. V. R. (2022). Effective Monitoring of Sensor-Actuator Integration Methods for Real-Time Controlling in Smart Device Networks. *International Journal of Emerging Research in Engineering and Technology*, *3*(1), 144-154.

Pilli, J., Jannu, P. K., Mohammad, J. A., Panchali, S. H., & Mylavarapu, K. B. (2022). Review of Supply Chain–Integrated CRM Systems for Business Agility. *International Journal of Emerging Research in Engineering and Technology*, *3*(1), 155-163.

Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., & Syed, M. H. (2022). Intelligent Predictive System of Cloud Resource Utilization Forecasting with Advanced Deep Learning. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(1), 143-151.

Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., & Manoharan, P. (2022). A Survey of Statistical Process Control and Its Applications in Manufacturing Quality Assurance. *International Journal of Emerging Trends in Computer Science and Information Technology*, *3*(1), 127-135.

Taluri, R. (2021). Cloud-Native Architectures for Enterprise Financial Data Management, Analytics, and Regulatory Reporting Compliance. International Journal of Emerging Trends in Computer Science and Information Technology, 2(2), 101-111. https://doi.org/10.63282/3050-9246.IJETCSIT-V2I2P112

Taluri, R. (2022). Cloud Data Engineering Strategies for Large-Scale Financial Data Integration and Intelligent Corporate Performance Reporting. International Journal of Emerging Research in Engineering and Technology, 3(4), 176-188. https://doi.org/10.63282/3050-922X.IJERET-V3I4P119