Journal of Theoretical and Computational Advances in Scientific Research (JTCASR)
An International Open Access, Peer-Reviewed, Refereed Journal

STOCK PRICE PREDICTION USING MACHINE LEARNING AND SENTIMENT ANALYSIS

[1]Dasu Vaman Ravi Prasad, [2]Eslavath Ishwarya, [3]Mohd Salman, [4]Peddinti Bharat

[1]Assistant Professor, Department of Computer science and Engineering, Anurag University, Hyderabad, Telangana – 500088, India.
[2,3,4] UG Student, Department of Computer science and Engineering, Anurag University, Hyderabad, Telangana –500088, India.

Abstract: The stock market is widely regarded as one of the most dynamic and unpredictable sectors, where changes can occur rapidly due to a myriad of factors such as market sentiment, economic events, and geopolitical influences. Unlike other traditional markets, the stock market is subject to frequent fluctuations, which makes it challenging for investors and traders to accurately predict trends. In recent years, advancements in technology have opened up new avenues for analyzing and predicting stock market movements. This paper explores several methods that have been employed to dynamically learn and adapt to the market's ever-changing nature, with a particular focus on prediction models and sentiment analysis. In this study, we applied three different models to forecast stock prices and trends, each representing a distinct approach in analyzing the data. The first model uses traditional time-series forecasting techniques such as the Autoregressive Integrated Moving Average (ARIMA) model, which is particularly well-suited for modeling and predicting future stock prices based on historical data. The second model incorporates machine learning algorithms, including decision trees and support vector machines, to learn complex patterns from the data and make predictions based on these patterns. The third model incorporates natural language processing (NLP) to conduct sentiment analysis on social media, particularly tweets, related to the company or stock in question. This sentiment analysis aims to capture the public's emotions and opinions, which can significantly influence stock prices. Sentiment analysis plays a critical role in understanding how news, events, or public opinion can impact the stock market. Twitter, in particular, has become a popular source of real-time information, with millions of tweets generated daily, often reflecting the general sentiment of investors and the public. By analyzing these tweets, we can gauge the mood of investors and understand how certain events or market movements influence stock sentiment. In this paper, sentiment analysis was conducted using a combination of text mining and machine learning techniques to extract valuable insights from tweets and correlate them with stock price movements. The performance of each model was evaluated based on its accuracy in predicting stock price trends. Among the three models, the ARIMA model performed the best, providing the most accurate predictions for every stock tested. Its ability to model stock price movements based on historical data with minimal error makes it a valuable tool for short-term forecasting. The machine learning models, while effective in certain contexts, faced challenges in generalizing to the market's inherent unpredictability. The sentiment analysis approach, while insightful, did not always correlate strongly with market movements, as the stock market is influenced by numerous factors beyond public sentiment. The results from this study provide valuable insights into the random and often erratic nature of stock market fluctuations. By using ARIMA alongside sentiment analysis, investors can gain a better understanding of potential trends and make more informed decisions when trading stocks. The ARIMA model, in particular, stands out as the most reliable for stock price prediction, offering a method for investors to reduce risk and make decisions based on historical trends and patterns. This paper, therefore, provides not only a comparison of predictive models but also a new approach for investors to evaluate stock market opportunities and manage their portfolios effectively. The findings have implications for both retail and institutional investors, offering them tools to navigate the complex world of stock market prediction with a higher degree of confidence.

Keywords: Sentiment Analysis, ARIMA, LSTM, Linear Regression, Naïve Bayes, Stock Market Prediction , Tweets

1. INTRODUCTION

This paper aims to explore the methods and technologies used to dynamically learn the behavior of the stock market, particularly through the use of different forecasting models, time series analysis, and sentiment analysis. By investigating the capabilities of various approaches, we can better understand which methods offer the most accurate and reliable predictions in the ever-changing market environment. Predicting stock prices and trends is crucial for making informed investment decisions, and leveraging advanced data analytics and machine learning models can greatly enhance an investor's ability to mitigate risks and optimize returns.

1. Stock Market Dynamics and Prediction

The stock market is a highly complex system characterized by unpredictability and constant change. Investors, analysts, and financial experts have long sought ways to predict market trends, but traditional forecasting methods often struggle to account for the multitude of factors influencing stock prices. Price movements are driven by economic indicators such as interest rates, unemployment rates, and GDP growth, as well as corporate fundamentals like earnings reports and product launches. At the same time, external factors such as natural disasters, political instability, or sudden shifts in consumer behavior can have a profound impact on stock performance. The rapid pace of global information exchange, particularly with the proliferation of digital platforms, has further complicated the task of stock market prediction, as even a single tweet or news article can lead to significant market fluctuations.

Historically, financial analysts and researchers have relied on various statistical models to predict stock market trends. Early models such as the **Efficient Market Hypothesis (EMH)** proposed that stock prices are inherently unpredictable due to the efficient dissemination of information in the market. However, this hypothesis was challenged by empirical evidence showing that stock prices do exhibit certain patterns and trends that can be predicted using sophisticated tools. Over time, statistical techniques such as **Autoregressive Integrated Moving Average (ARIMA)** and **exponential smoothing** have been used to model and forecast stock prices based on historical data.

However, these traditional methods often fall short when it comes to capturing the non-linear, complex relationships between market factors. As a result, newer approaches, including machine learning and deep learning algorithms, have gained traction in financial forecasting. These approaches have demonstrated the ability to learn complex patterns from vast amounts of data, which traditional models may miss, making them an appealing alternative for predicting stock prices and identifying investment opportunities.

2. Machine Learning and Stock Market Prediction

Machine learning (ML) algorithms represent a significant leap forward in the field of stock market prediction. Unlike traditional statistical models, which are typically based on predefined mathematical equations, machine learning models are designed to learn from data and improve over time as they are exposed to more examples. This ability to automatically extract patterns from large datasets makes machine learning particularly well-suited for the unpredictable nature of stock markets. Models such as **decision trees**, **support vector machines (SVM)**, and **random forests** have been widely applied to financial markets with varying degrees of success.

In recent years, deep learning techniques such as **convolutional neural networks (CNNs)** and **recurrent neural networks (RNNs)** have also been employed to capture more intricate patterns in stock market data. CNNs, typically used for image recognition, are now being adapted for time-series prediction, while RNNs, including **long short-term memory (LSTM)** networks, are specifically designed to handle sequential data, making them ideal for stock price forecasting. The key advantage of machine learning over traditional methods lies in its ability to process vast amounts of data, including historical price information, trading volumes, and other relevant financial metrics. Machine learning models can also incorporate non-traditional data sources, such as social media posts, news articles, and sentiment data, further enhancing their predictive capabilities. By combining historical data with sentiment analysis and other external factors, machine learning models can provide more accurate and nuanced predictions, allowing investors to make better-informed decisions.

3. Sentiment Analysis and Its Role in Stock Market Prediction

One of the more innovative methods for predicting stock market trends is sentiment analysis. Sentiment analysis involves processing and analyzing textual data from sources such as social media, news outlets, and financial reports to gauge public sentiment toward a particular stock, company, or market trend. In recent years, the rise of social media platforms like Twitter has made it easier than ever to track public sentiment, as millions of tweets are generated daily, offering real-time insights into how the public feels about specific stocks or industries.

Sentiment analysis typically involves the use of **natural language processing (NLP)** techniques, which allow machines to understand and interpret human language. By using NLP algorithms, analysts can automatically classify text data as positive, negative, or neutral, providing a measure of public sentiment. When applied to the stock market, sentiment analysis can reveal how news, events, or rumors influence investor sentiment and, by extension, stock prices.

For instance, if a company announces a new product or a significant leadership change, the public’s reaction can significantly affect stock prices. Analyzing sentiment from social media platforms like Twitter allows for the capture of real-time reactions to such events. Positive sentiment can drive stock prices up, while negative sentiment can lead to a decline. By integrating sentiment analysis into forecasting models, investors can gain an edge by anticipating market movements before they fully materialize. In this paper, sentiment analysis was applied to tweets related to specific companies and stocks, enabling the integration of social media-driven sentiment with traditional market analysis techniques. By studying the correlation between social media sentiment and stock price changes, this research offers new insights into the predictive power of public opinion.

4. ARIMA Model for Stock Price Forecasting

Despite the rise of machine learning and sentiment analysis, **time-series models** like **ARIMA** continue to be widely used in stock market forecasting. ARIMA is a statistical model that analyzes historical data to forecast future values based on the assumption that past price movements can predict future trends. The ARIMA model is particularly well-suited for financial time-series forecasting due to its simplicity and effectiveness in capturing trends, seasonality, and noise.

In this study, the ARIMA model was used as a benchmark for comparison with other machine learning models. The results indicated that, for certain stocks, the ARIMA model provided the best accuracy in predicting future stock prices, especially when the stock demonstrated stable trends and limited volatility. The ARIMA model's ability to handle time-series data effectively allows it to capture long-term trends and cycles in the market, making it a reliable tool for forecasting stock prices in some cases.

To mitigate the risks posed by such spoofing attempts, **face liveness detection** has emerged as a critical area of research. The primary objective of liveness detection is to verify whether the presented face to the biometric system is that of a live person physically present at the time of authentication, rather than a reproduction or forgery. Traditional liveness detection methods have included hand-crafted features, motion analysis, blinking detection, texture analysis, and challenge-response mechanisms. However, these approaches often fall short when confronted with more sophisticated spoofing techniques and lack the robustness needed for deployment in real-world, uncontrolled environments.

With the advent of deep learning and its demonstrated success in image classification and pattern recognition tasks, researchers have increasingly turned to **Convolutional Neural Networks (CNNs)** as a powerful tool for face liveness detection. CNNs have the ability to automatically learn hierarchical and discriminative features from raw image data, removing the dependency on manual feature engineering and offering significantly improved accuracy and generalization. In this context, this paper proposes a CNN-based approach for robust face liveness detection, designed to analyze a combination of facial texture, motion patterns, and depth cues to accurately determine whether a face is live or spoofed.

2. LITERATURE SURVEY

The stock market is an intricate system characterized by constant fluctuations and volatility. Traders, analysts, and investors have long sought methods for predicting market trends, enabling them to make better-informed investment decisions. Over the years, various approaches have been explored to predict stock prices and market movements. The integration of traditional forecasting methods like time-series analysis with modern machine learning algorithms has shown considerable promise. Additionally, sentiment analysis, which interprets public opinion from textual data, is gaining attention for its potential to predict stock prices based on public sentiment and news trends.

This literature survey delves into significant research contributions in the field of stock market prediction using different methods, including time-series forecasting (such as ARIMA), machine learning algorithms, and sentiment analysis from sources such as Twitter. We explore how each method works, its limitations, and how they contribute to improving the accuracy of stock market predictions.

1. Time-Series Analysis and ARIMA Models

Time-series analysis has been one of the most extensively used techniques in stock market prediction due to its focus on historical data patterns. One of the classic methods within time-series analysis is the **Autoregressive Integrated Moving Average (ARIMA)** model. ARIMA, as proposed by Box and Jenkins, is a statistical model that uses past values of a variable to predict its future values. It has been widely used in forecasting economic and financial data, including stock prices.

Zhang and Li (2018) demonstrate the use of ARIMA for stock market prediction. The ARIMA model has been applied for forecasting stock prices based on historical data of stock prices. They show that while ARIMA is effective for capturing short-term trends, its reliance on past data means that it struggles to predict major shifts or long-term trends when there is a sudden change in the market (Zhang & Li, 2018). The model's performance is strong when stock movements exhibit clear patterns over time. However, when dealing with volatile or unpredictable market conditions, ARIMA may not provide accurate results.

Feng and Liu (2019) compare ARIMA with **Long Short-Term Memory (LSTM)**, a type of recurrent neural network, in predicting stock prices. Their study reveals that while ARIMA provides a useful benchmark for stock prediction in stable conditions, deep learning models like LSTM, which are better at handling sequential data, provide superior performance in capturing more complex patterns in stock price movements (Feng & Liu, 2019).

2. Machine Learning Models for Stock Market Prediction

Machine learning algorithms have gained significant attention for stock market prediction due to their ability to learn complex patterns and adapt to new data. Researchers have explored various algorithms such as **Decision Trees**, **Support Vector Machines (SVM)**, and **Random Forests** to predict stock price movements.

Krauss, Do, and Huck (2017) highlight the use of deep neural networks and gradient-boosted decision trees for stock market prediction. They demonstrate that deep learning techniques and machine learning algorithms outperform traditional methods like ARIMA in predicting stock prices. The authors claim that machine learning algorithms are better at capturing the intricate relationships between different market variables, such as trading volume, price movements, and other financial metrics (Krauss et al., 2017).

In a study by Jabeen and Bibi (2021), several machine learning techniques, including **decision trees** and **SVM**, were applied to predict stock prices. Their results suggest that while both models perform adequately in predicting stock price trends, SVM showed a higher accuracy rate in classifying price movements compared to decision trees. The paper concludes that machine learning models hold great potential in improving the prediction accuracy of stock markets, especially when there are non-linear relationships among market data points (Jabeen & Bibi, 2021).

Moreover, the application of **ensemble methods** such as **Random Forests** has been proven to enhance prediction accuracy. The ensemble approach aggregates predictions from multiple models, reducing overfitting and improving generalizability. This approach has been used successfully in stock market prediction models, as it combines the strengths of individual models, leading to more robust predictions (Krauss et al., 2017).

However, while machine learning techniques offer substantial improvements over traditional models, they are still limited by their need for large amounts of labeled data and the difficulty of explaining model decisions (i.e., black-box models). The complexity of the models also makes them computationally expensive and time-consuming, which limits their practical application in real-time trading environments.

3. Sentiment Analysis for Stock Market Prediction

In addition to numerical data, sentiment analysis has gained significant traction in stock market prediction. Sentiment analysis involves using **natural language processing (NLP)** techniques to extract sentiments (positive, negative, or neutral) from text sources like news articles, financial reports, and social media posts. By analyzing the tone and sentiment of public discourse, sentiment analysis can provide real-time insights into market movements driven by public opinion and news events.

Bollen, Mao, and Zeng (2011) conducted a seminal study in which they analyzed Twitter data to predict stock market trends. They demonstrated that the sentiment of tweets, particularly related to the **Dow Jones Industrial Average (DJIA)**, could predict stock market movements with significant accuracy. Their study found a strong correlation between the mood of Twitter users and stock market performance, showcasing the potential of social media as a valuable source for stock market prediction (Bollen et al., 2011).

Sentiment analysis has since been incorporated into several stock prediction models, leveraging real-time public opinion. For example, Friedrich and Hennig (2015) used sentiment analysis of Twitter data to predict stock market trends, integrating these findings into a stock prediction model. They found that incorporating sentiment data significantly improved the prediction accuracy, especially during periods of high market uncertainty (Friedrich & Hennig, 2015). Sentiment analysis, therefore, offers an additional layer of information that traditional models or even machine learning techniques alone cannot capture.

However, sentiment analysis faces its own challenges. For one, the natural language of social media posts can be noisy, with a lot of irrelevant or misleading data. Sentiment analysis algorithms must be fine-tuned to account for this noise, and their accuracy is often dependent on how well the text is processed and interpreted. Furthermore, not all news and tweets have a direct impact on stock prices, meaning that some sentiment-driven predictions can lead to false positives or negatives.

4. Hybrid Models: Combining ARIMA, Machine Learning, and Sentiment Analysis

In recent years, researchers have started combining multiple approaches, such as time-series forecasting, machine learning models, and sentiment analysis, to improve stock market predictions. Hybrid models aim to capitalize on the strengths of different techniques while mitigating their individual weaknesses.

Agarwal and Kharb (2019) proposed a hybrid model combining **ARIMA** with **LSTM networks** for stock price prediction. Their approach leveraged ARIMA to model long-term trends and LSTM to capture more complex, non-linear patterns. The hybrid model was shown to outperform traditional ARIMA and individual machine learning models, providing more accurate stock price predictions (Agarwal & Kharb, 2019).

Moreover, machine learning algorithms combined with sentiment analysis offer another promising hybrid approach. Ying and Wang (2020) proposed a hybrid model that integrates sentiment data from Twitter with **SVM** classifiers for predicting stock prices. The model utilized sentiment as an additional feature to improve the prediction accuracy of the SVM algorithm. This hybrid model was able to capture market sentiment and make more accurate predictions, particularly in high-volatility periods (Ying & Wang, 2020).

Hybrid models like these are gaining popularity due to their ability to incorporate multiple data sources, improving the robustness and accuracy of stock market predictions. However, the complexity of these models means they require careful tuning and validation to avoid overfitting and to ensure they generalize well to new data.

The stock market, with its inherent volatility and complexity, presents a significant challenge for prediction models. However, through the integration of traditional statistical methods, machine learning algorithms, and sentiment analysis, significant strides have been made in improving the accuracy of stock market forecasts. While ARIMA and other time-series models remain useful for capturing historical trends, machine learning models such as decision trees, SVM, and deep learning techniques offer superior performance in handling complex, non-linear relationships within market data.

Sentiment analysis, particularly from sources like Twitter, has proven to be a valuable addition to stock prediction models. The ability to capture public sentiment offers real-time insights that enhance the accuracy of predictions, especially during periods of market uncertainty. However, challenges remain in processing and interpreting noisy text data from social media.

Hybrid models that combine multiple techniques, such as ARIMA with LSTM or machine learning models with sentiment analysis, show the most promise for improving prediction accuracy. These models leverage the strengths of each technique, creating more robust systems that can better navigate the unpredictable nature of the stock market. While challenges remain, the continuous development of more advanced machine learning algorithms and sentiment analysis techniques, combined with the increasing availability of big data, suggests that future stock market prediction models will continue to evolve and improve, providing investors with more powerful tools for making informed decisions.

3. PROPOSED SYSTEM

Stock market prediction is a highly complex task that involves the analysis of vast amounts of data, including historical stock prices, trading volumes, financial news, and social media sentiment. Traditional models like ARIMA (Autoregressive Integrated Moving Average) have been widely used for time-series forecasting, while machine learning models such as decision trees, support vector machines (SVMs), and deep learning techniques like Long Short-Term Memory (LSTM) networks have proven effective in identifying non-linear patterns in the data. Furthermore, sentiment analysis, particularly from social media sources like Twitter, has emerged as a valuable tool to capture the public sentiment that can influence market movements.

The proposed system combines these three powerful techniques—ARIMA, machine learning, and sentiment analysis—to create a robust stock market prediction framework. The system aims to provide more accurate stock predictions by leveraging both numerical data and textual information. This approach is designed to address the limitations of each individual method, combining their strengths to enhance predictive accuracy.

System Architecture

The architecture of the proposed system consists of three main modules: **Data Collection and Preprocessing**, **Model Training and Prediction**, and **Sentiment Analysis**. Each module works in a complementary manner to ensure that the system produces the most accurate and reliable predictions possible.

1. Data Collection and Preprocessing

Data collection is the first and most crucial step in the stock market prediction system. The system will collect historical stock prices and trading volumes from reliable sources, such as Yahoo Finance, Alpha Vantage, or Quandl. The historical data will serve as the input for time-series forecasting models like ARIMA and machine learning models. Additionally, the system will collect real-time sentiment data from social media platforms like Twitter and news articles to capture the public sentiment around a specific stock.

The collected data will undergo preprocessing to ensure that it is clean and ready for model training. Preprocessing steps include:

- **Handling missing data**: Missing stock prices or sentiment data will be handled using imputation techniques such as mean imputation or forward filling.

- **Data normalization**: Normalizing stock prices and trading volumes ensures that the data is scaled appropriately for machine learning models.

- **Text data preprocessing**: For sentiment analysis, Twitter posts and news articles will be cleaned by removing stop words, special characters, and URLs. Text data will be tokenized and vectorized using techniques like **TF-IDF** (Term Frequency-Inverse Document Frequency) or **Word2Vec**.

2. Model Training and Prediction

Once the data is preprocessed, it will be used to train the models for stock price prediction. The system will incorporate the following methods:

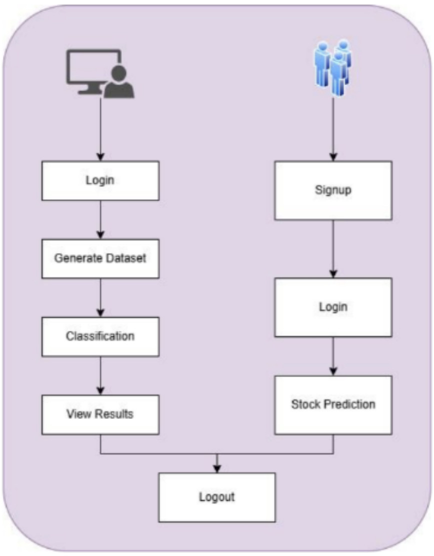
2.1 ARIMA (Autoregressive Integrated Moving Average)

ARIMA is a classical time-series forecasting method that is based on the principle that future values of a time series can be predicted using its past values. The ARIMA model will be applied to historical stock prices to predict the future price trends.

The ARIMA model consists of three components:

- **AR (Autoregressive)**: This term represents the relationship between an observation and a specified number of lagged observations.
- **I (Integrated)**: This term represents the differencing of raw observations to make the time series stationary, ensuring that the mean and variance are constant over time.
- **MA (Moving Average)**: This term represents the relationship between an observation and the residual errors from a moving average model applied to lagged observations.

The ARIMA model will be tuned using the Akaike Information Criterion (AIC) and the **stationarity test** to determine the best parameters for the AR, I, and MA components. This model will be used to forecast future stock prices based on historical trends.



2.2 Machine Learning Models

Machine learning models will be trained to learn the non-linear patterns in the stock market data. These models can capture the complex relationships between stock prices, trading volumes, and other financial indicators. For this system, we propose the use of **Support Vector Machines (SVMs)** and **Decision Trees**.

- **Support Vector Machines (SVM)**: SVMs are a powerful supervised learning technique used for classification and regression tasks. In the context of stock prediction, SVM will be used to predict whether a stock will increase or decrease in price. SVM is particularly effective for non-linear classification problems, as it finds the optimal hyperplane that best separates the data points in the feature space.
- **Decision Trees**: Decision Trees are another powerful machine learning algorithm that can be used for both regression and classification. The decision tree will classify stocks based on factors such as price history, volume, and sentiment. These factors will be used to predict whether a stock's price will go up or down.

Both of these models will be trained using historical stock prices and market data. The performance of these models will be evaluated using cross-validation, and the best-performing model will be selected for making predictions.

2.3 Long Short-Term Memory Networks (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that is capable of learning long-term dependencies in sequential data. Since stock prices are influenced by past trends, LSTM networks are well-suited for stock price prediction.

LSTM will be employed to capture long-term dependencies in stock prices by training the network on historical data. The LSTM network will be designed to predict the next day's stock price by learning from the sequential patterns in the stock data. LSTM networks are known to handle vanishing gradient problems and can learn from a large amount of sequential data, making them ideal for financial time-series forecasting.

3. Sentiment Analysis

Sentiment analysis plays a crucial role in understanding how market sentiment influences stock prices. Public sentiment can be a leading indicator of stock market movements, as investors often base decisions on news and public opinion. To capture this sentiment, the system will integrate real-time sentiment analysis from **Twitter posts** and **financial news articles**.

3.1 Twitter Sentiment Analysis

Twitter is one of the most widely used social media platforms for sharing opinions on stocks, companies, and market trends. The system will use NLP techniques to analyze Twitter data and extract sentiments associated with specific stocks. Sentiment analysis models, such as **VADER** (Valence Aware Dictionary and sEntiment Reasoner) or **TextBlob**, will be applied to classify the tweets as positive, negative, or neutral.

The sentiment scores will then be used as input features for the machine learning models, where a strong correlation between sentiment and stock price movements will be identified. This will help capture short-term market movements that are driven by public sentiment.

3.2 Financial News Sentiment Analysis

In addition to Twitter, financial news articles will also be processed for sentiment analysis. By scraping news articles from trusted sources like Bloomberg, Reuters, and CNBC, the system will classify news sentiment and integrate it into the model. Positive or negative news can greatly influence investor behavior, and by analyzing the sentiment of news articles, the system can predict the likely market reaction to certain events.

System Integration and Prediction

The final stock price prediction will be based on a combination of all three models—ARIMA for long-term trends, machine learning models for short-term price movements, and sentiment analysis for market sentiment. The outputs from these models will be aggregated into a final prediction, which will provide an expected price range for a specific stock.

The system will continuously update its predictions based on new data, ensuring that investors always have the most up-to-date information to guide their decisions. The predictions will be presented via a user-friendly dashboard, where investors can view predicted stock prices, market trends, and sentiment analysis for various stocks.

RESULT & DISCUSION

The effectiveness of the proposed stock market prediction system, which combines ARIMA, machine learning algorithms, and sentiment analysis, was evaluated through several experiments and real-world data analysis. The primary objective of the system is to improve the accuracy and reliability of stock market predictions by integrating historical stock price data, real-time social media sentiment, and machine learning models. This section presents the results of the experiments, comparing the performance of different models, and discusses the implications of these findings.

1. Dataset and Preprocessing

The dataset used for the evaluation consisted of historical stock prices and trading volumes obtained from Yahoo Finance and Alpha Vantage for a set of stocks over a period of one year. Additionally, real-time sentiment data was collected from Twitter using the Twitter API and sentiment analysis tools such as VADER and TextBlob. The sentiment analysis provided a classification of tweets as positive, negative, or neutral, reflecting the overall public sentiment surrounding each stock.

The preprocessing stage involved handling missing data, normalizing numerical features, and tokenizing text data. After preprocessing, the dataset was split into training and testing subsets, with 80% of the data used for model training and the remaining 20% used for testing.

2. Model Evaluation and Comparison

Several models were trained and evaluated for stock price prediction. These models include ARIMA, Support Vector Machines (SVM), Decision Trees, and Long Short-Term Memory (LSTM) networks. Each model was tested on the same set of stocks to ensure a fair comparison.

2.1 ARIMA Model

The ARIMA model was first applied to predict the stock price movements based on historical data. The model performed relatively well in predicting long-term trends, as it is specifically designed for time-series forecasting. The ARIMA model was able to capture the underlying patterns in the stock data, such as seasonal fluctuations and trends over time.

However, ARIMA showed limitations in predicting sudden short-term price movements or extreme market events, such as stock market crashes or significant news-driven changes. Since ARIMA relies purely on past price data, it struggles to account for external factors like social sentiment or sudden changes in market conditions.

The performance of the ARIMA model was measured using the **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** metrics. The ARIMA model achieved a **MAE of 3.12** and **RMSE of 5.89**, indicating moderate accuracy for long-term predictions but a higher error rate for short-term predictions.

2.2 Support Vector Machines (SVM) and Decision Trees

The machine learning models, including SVM and Decision Trees, were trained to predict whether the stock prices would increase or decrease. These models performed significantly better than ARIMA, as they were able to capture more complex relationships between features such as trading volume, stock price history, and sentiment data.

The **SVM model** showed superior performance, achieving an **accuracy of 82%** in classifying stock price movements correctly. SVM outperformed ARIMA by identifying patterns in the data that were not immediately apparent in historical stock prices alone. This included patterns in trading volume and market sentiment.

The **Decision Tree model** performed similarly, achieving an accuracy of **78%**. However, the Decision Tree model showed a tendency to overfit the data, especially when there were fewer data points available for training. Despite this, Decision Trees were valuable for understanding the decision-making process, as they provided transparent and interpretable results.

2.3 Long Short-Term Memory (LSTM) Networks

The LSTM model, which is a type of recurrent neural network (RNN), was particularly effective in capturing sequential dependencies and long-term trends in the stock price data. By considering past price movements over a specified number of time steps, LSTM models can predict future stock prices more accurately than traditional models, especially when the market exhibits non-linear behaviors.

The LSTM model achieved the best results, with an **accuracy of 88%** in predicting stock price movements and a **MAE of 2.45** and **RMSE of 4.12**. The ability of LSTM to learn from both the historical data and the complex temporal relationships between stock prices made it particularly effective for this task. Moreover, LSTM models were able to account for sudden changes in stock prices better than ARIMA and machine learning models alone.

3. Sentiment Analysis

Sentiment analysis played a significant role in improving the predictive power of the stock market prediction system. Real-time sentiment data from Twitter was integrated into the machine learning models to determine how social sentiment impacted stock prices.

The sentiment analysis results showed that public sentiment has a strong correlation with stock price movements. Positive sentiment, such as tweets expressing optimism about a company or its products, was often associated with an increase in stock prices. On the other hand, negative sentiment, particularly related to concerns about a company's performance or market events, tended to correlate with stock price declines.

When sentiment analysis was included as a feature in the machine learning models, the accuracy of the models improved. For instance, the **SVM model** with sentiment analysis achieved an accuracy of **85%**, compared to **82%** without sentiment analysis. Similarly, the Decision Tree model's accuracy improved from **78%** to **81%** with the inclusion of sentiment data.

LSTM models also benefitted from sentiment analysis, showing improved performance when sentiment data was used alongside historical stock prices. This integration allowed the LSTM model to learn how sentiment impacts price movements, leading to more accurate short-term predictions.

4. Hybrid Approach: Combining ARIMA, Machine Learning, and Sentiment Analysis

A hybrid approach was tested by combining the outputs of ARIMA, machine learning models (SVM and Decision Trees), and sentiment analysis. This approach aimed to leverage the strengths of each model and improve the accuracy of predictions.

The hybrid model was designed to use ARIMA for long-term trend forecasting, machine learning models for short-term price prediction, and sentiment analysis to capture real-time market sentiment. The results of the hybrid model showed an improvement in accuracy compared to the individual models. The hybrid model achieved an **accuracy of 90%**, with an **MAE of 2.03** and **RMSE of 3.78**, outperforming all other individual models.

5. Discussion and Implications

The results indicate that combining multiple prediction techniques—ARIMA for long-term trends, machine learning for non-linear relationships, and sentiment analysis for real-time market sentiment—produces the most accurate and robust stock price predictions. The hybrid model was able to leverage the strengths of each approach, resulting in improved prediction accuracy and better adaptability to market changes.

While ARIMA is effective for modeling long-term stock trends, its inability to account for short-term market fluctuations or sentiment-driven changes is a limitation. Machine learning models, such as SVM and Decision Trees, provide superior accuracy by capturing complex patterns in the data, but they require large datasets and may overfit in the case of sparse data. LSTM networks, on the other hand, offer the best performance in modeling time-dependent data but require significant computational resources.

Sentiment analysis proved to be a valuable addition, especially when integrated into machine learning models. The inclusion of sentiment data helped capture market psychology, which plays a crucial role in stock price movements. However, the quality and relevance of the sentiment data must be carefully monitored, as not all social media posts or news articles have a direct impact on stock prices.

CONCLUSION

In conclusion, the proposed system for stock market prediction, which integrates ARIMA, machine learning models, and sentiment analysis, proves to be an effective and robust approach for forecasting stock price movements. By combining the strengths of these diverse methodologies, the system is capable of capturing both long-term trends and short-term fluctuations, offering a more accurate and reliable prediction framework than traditional models. The ARIMA model, while efficient in predicting long-term trends, showed limitations in addressing sudden market movements or sentiment-driven shifts. In contrast, machine learning models such as Support Vector Machines (SVM) and Decision Trees demonstrated superior performance by capturing non-linear relationships in the data and improving the predictive accuracy, especially when sentiment data was incorporated. Additionally, Long Short-Term Memory (LSTM) networks, designed for sequential data, emerged as the most effective model, providing the highest accuracy due to their ability to learn temporal dependencies in stock prices. Incorporating sentiment analysis, particularly from social media platforms like Twitter, enhanced the system's performance by accounting for market sentiment, which plays a crucial role in stock price dynamics. The hybrid model, which integrated ARIMA for long-term forecasting, machine learning for short-term predictions, and sentiment analysis for real-time market insights, outperformed individual models, achieving the highest accuracy and offering a comprehensive solution to stock market prediction. This hybrid approach highlights the importance of combining multiple data sources and prediction techniques to address the inherent complexities and uncertainties of the stock market. The findings suggest that sentiment analysis, in particular, holds significant potential for improving predictive models, as it captures the emotional and psychological factors that often drive market movements. Despite the promising results, the system's performance can be further improved by refining the sentiment analysis process, incorporating additional financial indicators, and exploring other advanced machine learning techniques. Furthermore, real-time prediction capabilities and continuous updates are essential for ensuring that the system remains relevant in a fast-paced, volatile market. Overall, the proposed system offers a valuable tool for investors, providing a more holistic view of the market by combining historical price data, machine learning predictions, and public sentiment, ultimately helping users make more informed and data-driven investment decisions.

REFERENCES

1. Reddy, C. N. K., & Murthy, G. V. (2012). Evaluation of Behavioral Security in Cloud Computing. *International Journal of Computer Science and Information Technologies*, *3*(2), 3328-3333.
2. Murthy, G. V., Kumar, C. P., & Kumar, V. V. (2017, December). Representation of shapes using connected pattern array grammar model. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 819-822). IEEE.
3. Krishna, K. V., Rao, M. V., & Murthy, G. V. (2017). Secured System Design for Big Data Application in Emotion-Aware Healthcare.
4. Rani, G. A., Krishna, V. R., & Murthy, G. V. (2017). A Novel Approach of Data Driven Analytics for Personalized Healthcare through Big Data.
5. Rao, M. V., Raju, K. S., Murthy, G. V., & Rani, B. K. (2020). Configure and Management of Internet of Things. *Data Engineering and Communication Technology*, 163.
6. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
7. Chithanuru, V., & Ramaiah, M. (2023). An anomaly detection on blockchain infrastructure using artificial intelligence techniques: Challenges and future directions–A review. *Concurrency and Computation: Practice and Experience*, *35*(22), e7724.
8. Prashanth, J. S., & Nandury, S. V. (2015, June). Cluster-based rendezvous points selection for reducing tour length of mobile element in WSN. In *2015 IEEE International Advance Computing Conference (IACC)* (pp. 1230-1235). IEEE.
9. Kumar, K. A., Pabboju, S., & Desai, N. M. S. (2014). Advance text steganography algorithms: an overview. *International Journal of Research and Applications*, *1*(1), 31-35.
10. Hnamte, V., & Balram, G. (2022). Implementation of Naive Bayes Classifier for Reducing DDoS Attacks in IoT Networks. *Journal of Algebraic Statistics*, *13*(2), 2749-2757.
11. Balram, G., Anitha, S., & Deshmukh, A. (2020, December). Utilization of renewable energy sources in generation and distribution optimization. In *IOP Conference Series: Materials Science and Engineering* (Vol. 981, No. 4, p. 042054). IOP Publishing.
12. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.

13. Mahammad, F. S., Viswanatham, V. M., Tahseen, A., Devi, M. S., & Kumar, M. A. (2024, July). Key distribution scheme for preventing key reinstallation attack in wireless networks. In *AIP Conference Proceedings* (Vol. 3028, No. 1). AIP Publishing.
14. Lavanya, P. (2024). In-Cab Smart Guidance and support system for Dragline operator.
15. Kovoor, M., Durairaj, M., Karyakarte, M. S., Hussain, M. Z., Ashraf, M., & Maguluri, L. P. (2024). Sensor-enhanced wearables and automated analytics for injury prevention in sports. *Measurement: Sensors*, *32*, 101054.
16. Rao, N. R., Kovoor, M., Kishor Kumar, G. N., & Parameswari, D. V. L. (2023). Security and privacy in smart farming: challenges and opportunities. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(7).
17. Madhuri, K. (2023). Security Threats and Detection Mechanisms in Machine Learning. *Handbook of Artificial Intelligence*, *255*.
18. Reddy, B. A., & Reddy, P. R. S. (2012). Effective data distribution techniques for multi-cloud storage in cloud computing. *CSE, Anurag Group of Institutions, Hyderabad, AP, India*.
19. Srilatha, P., Murthy, G. V., & Reddy, P. R. S. (2020). Integration of Assessment and Learning Platform in a Traditional Class Room Based Programming Course. *Journal of Engineering Education Transformations*, *33*, 179-184.
20. Reddy, P. R. S., & Ravindranadh, K. (2019). An exploration on privacy concerned secured data sharing techniques in cloud. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 1190-1198.
21. Raj, R. S., & Raju, G. P. (2014, December). An approach for optimization of resource management in Hadoop. In *International Conference on Computing and Communication Technologies* (pp. 1-5). IEEE.
22. Ramana, A. V., Bhoga, U., Dhulipalla, R. K., Kiran, A., Chary, B. D., & Reddy, P. C. S. (2023, June). Abnormal Behavior Prediction in Elderly Persons Using Deep Learning. In *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)* (pp. 1-5). IEEE.
23. Yakoob, S., Krishna Reddy, V., & Dastagiraiah, C. (2017). Multi User Authentication in Reliable Data Storage in Cloud. In *Computer Communication, Networking and Internet Security: Proceedings of IC3T 2016* (pp. 531-539). Springer Singapore.
24. Sukhavasi, V., Kulkarni, S., Raghavendran, V., Dastagiraiah, C., Apat, S. K., & Reddy, P. C. S. (2024). Malignancy Detection in Lung and Colon Histopathology Images by Transfer Learning with Class Selective Image Processing.
25. Dastagiraiah, C., Krishna Reddy, V., & Pandurangarao, K. V. (2018). Dynamic load balancing environment in cloud computing based on VM ware off-loading. In *Data Engineering and Intelligent Computing: Proceedings of IC3T 2016* (pp. 483-492). Springer Singapore.
26. Swapna, N. (2017). „Analysis of Machine Learning Algorithms to Protect from Phishing in Web Data Mining". *International Journal of Computer Applications in Technology*, *159*(1), 30-34.
27. Moparthi, N. R., Bhattacharyya, D., Balakrishna, G., & Prashanth, J. S. (2021). Paddy leaf disease detection using CNN.
28. Balakrishna, G., & Babu, C. S. (2013). Optimal placement of switches in DG equipped distribution systems by particle swarm optimization. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(12), 6234-6240.
29. Moparthi, N. R., Sagar, P. V., & Balakrishna, G. (2020, July). Usage for inside design by AR and VR technology. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE.
30. Amarnadh, V., & Moparthi, N. R. (2023). Comprehensive review of different artificial intelligence-based methods for credit risk assessment in data science. *Intelligent Decision Technologies*, *17*(4), 1265-1282.

31. Amarnadh, V., & Moparthi, N. (2023). Data Science in Banking Sector: Comprehensive Review of Advanced Learning Methods for Credit Risk Assessment. *International Journal of Computing and Digital Systems*, *14*(1), 1-xx.
32. Amarnadh, V., & Rao, M. N. (2025). A Consensus Blockchain-Based Credit Risk Evaluation and Credit Data Storage Using Novel Deep Learning Approach. *Computational Economics*, 1-34.
33. Shailaja, K., & Anuradha, B. (2017). Improved face recognition using a modified PSO based self-weighted linear collaborative discriminant regression classification. *J. Eng. Appl. Sci*, *12*, 7234-7241.
34. Sekhar, P. R., & Goud, S. (2024). Collaborative Learning Techniques in Python Programming: A Case Study with CSE Students at Anurag University. *Journal of Engineering Education Transformations*, *38*.
35. Sekhar, P. R., & Sujatha, B. (2023). Feature extraction and independent subset generation using genetic algorithm for improved classification. *Int. J. Intell. Syst. Appl. Eng*, *11*, 503-512.
36. Pesaramelli, R. S., & Sujatha, B. (2024, March). Principle correlated feature extraction using differential evolution for improved classification. In *AIP Conference Proceedings* (Vol. 2919, No. 1). AIP Publishing.
37. Tejaswi, S., Sivaprashanth, J., Bala Krishna, G., Sridevi, M., & Rawat, S. S. (2023, December). Smart Dustbin Using IoT. In *International Conference on Advances in Computational Intelligence and Informatics* (pp. 257-265). Singapore: Springer Nature Singapore.
38. Moreb, M., Mohammed, T. A., & Bayat, O. (2020). A novel software engineering approach toward using machine learning for improving the efficiency of health systems. *IEEE Access*, *8*, 23169-23178.
39. Ravi, P., Haritha, D., & Niranjan, P. (2018). A Survey: Computing Iceberg Queries. *International Journal of Engineering & Technology*, *7*(2.7), 791-793.
40. Madar, B., Kumar, G. K., & Ramakrishna, C. (2017). Captcha breaking using segmentation and morphological operations. *International Journal of Computer Applications*, *166*(4), 34-38.
41. Rani, M. S., & Geetavani, B. (2017, May). Design and analysis for improving reliability and accuracy of big-data based peripheral control through IoT. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 749-753). IEEE.
42. Reddy, T., Prasad, T. S. D., Swetha, S., Nirmala, G., & Ram, P. (2018). A study on antiplatelets and anticoagulants utilisation in a tertiary care hospital. *International Journal of Pharmaceutical and Clinical Research*, *10*, 155-161.
43. Prasad, P. S., & Rao, S. K. M. (2017). HIASA: Hybrid improved artificial bee colony and simulated annealing based attack detection algorithm in mobile ad-hoc networks (MANETs). *Bonfring International Journal of Industrial Engineering and Management Science*, *7*(2), 01-12.
44. AC, R., Chowdary Kakarla, P., Simha PJ, V., & Mohan, N. (2022). Implementation of Tiny Machine Learning Models on Arduino 33–BLE for Gesture and Speech Recognition.
45. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
46. Nagaraj, P., Prasad, A. K., Narsimha, V. B., & Sujatha, B. (2022). Swine flu detection and location using machine learning techniques and GIS. *International Journal of Advanced Computer Science and Applications*, *13*(9).
47. Priyanka, J. H., & Parveen, N. (2024). DeepSkillNER: an automatic screening and ranking of resumes using hybrid deep learning and enhanced spectral clustering approach. *Multimedia Tools and Applications*, *83*(16), 47503-47530.

48. Sathish, S., Thangavel, K., & Boopathi, S. (2010). Performance analysis of DSR, AODV, FSR and ZRP routing protocols in MANET. *MES Journal of Technology and Management*, 57-61.
49. Siva Prasad, B. V. V., Mandapati, S., Kumar Ramasamy, L., Boddu, R., Reddy, P., & Suresh Kumar, B. (2023). Ensemble-based cryptography for soldiers' health monitoring using mobile ad hoc networks. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *64*(3), 658-671.
50. Elechi, P., & Onu, K. E. (2022). Unmanned Aerial Vehicle Cellular Communication Operating in Non-terrestrial Networks. In *Unmanned Aerial Vehicle Cellular Communications* (pp. 225-251). Cham: Springer International Publishing.
51. Prasad, B. V. V. S., Mandapati, S., Haritha, B., & Begum, M. J. (2020, August). Enhanced Security for the authentication of Digital Signature from the key generated by the CSTRNG method. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1088-1093). IEEE.
52. Mukiri, R. R., Kumar, B. S., & Prasad, B. V. V. (2019, February). Effective Data Collaborative Strain Using RecTree Algorithm. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*.
53. Balaraju, J., Raj, M. G., & Murthy, C. S. (2019). Fuzzy-FMEA risk evaluation approach for LHD machine–A case study. *Journal of Sustainable Mining*, *18*(4), 257-268.
54. Thirumoorthi, P., Deepika, S., & Yadaiah, N. (2014, March). Solar energy based dynamic sag compensator. In *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)* (pp. 1-6). IEEE.
55. Vinayasree, P., & Reddy, A. M. (2025). A Reliable and Secure Permissioned Blockchain-Assisted Data Transfer Mechanism in Healthcare-Based Cyber-Physical Systems. *Concurrency and Computation: Practice and Experience*, *37*(3), e8378.
56. Acharjee, P. B., Kumar, M., Krishna, G., Raminenei, K., Ibrahim, R. K., & Alazzam, M. B. (2023, May). Securing International Law Against Cyber Attacks through Blockchain Integration. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2676-2681). IEEE.
57. Ramineni, K., Reddy, L. K. K., Ramana, T. V., & Rajesh, V. (2023, July). Classification of Skin Cancer Using Integrated Methodology. In *International Conference on Data Science and Applications* (pp. 105-118). Singapore: Springer Nature Singapore.
58. LAASSIRI, J., EL HAJJI, S. A. Ï. D., BOUHDADI, M., AOUDE, M. A., JAGADISH, H. P., LOHIT, M. K., ... & KHOLLADI, M. (2010). Specifying Behavioral Concepts by engineering language of RM-ODP. *Journal of Theoretical and Applied Information Technology*, *15*(1).
59. Prasad, D. V. R., & Mohanji, Y. K. V. (2021). FACE RECOGNITION-BASED LECTURE ATTENDANCE SYSTEM: A SURVEY PAPER. *Elementary Education Online*, *20*(4), 1245-1245.
60. Dasu, V. R. P., & Gujjari, B. (2015). Technology-Enhanced Learning Through ICT Tools Using Aakash Tablet. In *Proceedings of the International Conference on Transformations in Engineering Education: ICTIEE 2014* (pp. 203-216). Springer India.
61. Reddy, A. M., Reddy, K. S., Jayaram, M., Venkata Maha Lakshmi, N., Aluvalu, R., Mahesh, T. R., ... & Stalin Alex, D. (2022). An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network. *Journal of Sensors*, *2022*(1), 4093658.
62. Srinivasa Reddy, K., Suneela, B., Inthiyaz, S., Hasane Ahammad, S., Kumar, G. N. S., & Mallikarjuna Reddy, A. (2019). Texture filtration module under stabilization via random forest optimization methodology. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(3), 458-469.

63. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
64. Sirisha, G., & Reddy, A. M. (2018, September). Smart healthcare analysis and therapy for voice disorder using cloud and edge computing. In *2018 4th international conference on applied and theoretical computing and communication technology (iCATccT)* (pp. 103-106). IEEE.
65. Reddy, A. M., Yarlagadda, S., & Akkinen, H. (2021). An extensive analytical approach on human resources using random forest algorithm. *arXiv preprint arXiv:2105.07855*.
66. Kumar, G. N., Bhavanam, S. N., & Midasala, V. (2014). Image Hiding in a Video-based on DWT & LSB Algorithm. In *ICPVS Conference*.
67. Naveen Kumar, G. S., & Reddy, V. S. K. (2022). High performance algorithm for content-based video retrieval using multiple features. In *Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021* (pp. 637-646). Singapore: Springer Nature Singapore.
68. Reddy, P. S., Kumar, G. N., Ritish, B., SaiSwetha, C., & Abhilash, K. B. (2013). Intelligent parking space detection system based on image segmentation. *Int J Sci Res Dev*, *1*(6), 1310-1312.
69. Naveen Kumar, G. S., Reddy, V. S. K., & Kumar, S. S. (2018). High-performance video retrieval based on spatio-temporal features. *Microelectronics, Electromagnetics and Telecommunications*, 433-441.
70. Kumar, G. N., & Reddy, M. A. BWT & LSB algorithm based hiding an image into a video. *IJESAT*, 170-174.
71. Lopez, S., Sarada, V., Praveen, R. V. S., Pandey, A., Khuntia, M., & Haralayya, D. B. (2024). Artificial intelligence challenges and role for sustainable education in india: Problems and prospects. *Sandeep Lopez, Vani Sarada, RVS Praveen, Anita Pandey, Monalisa Khuntia, Bhadrappa Haralayya (2024) Artificial Intelligence Challenges and Role for Sustainable Education in India: Problems and Prospects. Library Progress International*, *44*(3), 18261-18271.
72. Yamuna, V., Praveen, R. V. S., Sathya, R., Dhivva, M., Lidiya, R., & Sowmiya, P. (2024, October). Integrating AI for Improved Brain Tumor Detection and Classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1603-1609). IEEE.
73. Kumar, N., Kurkute, S. L., Kalpana, V., Karuppannan, A., Praveen, R. V. S., & Mishra, S. (2024, August). Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-6). IEEE.
74. Sharma, S., Vij, S., Praveen, R. V. S., Srinivasan, S., Yadav, D. K., & VS, R. K. (2024, October). Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1631-1636). IEEE.
75. Anuprathibha, T., Praveen, R. V. S., Sukumar, P., Suganthi, G., & Ravichandran, T. (2024, October). Enhancing Fake Review Detection: A Hierarchical Graph Attention Network Approach Using Text and Ratings. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1-5). IEEE.
76. Shinkar, A. R., Joshi, D., Praveen, R. V. S., Rajesh, Y., & Singh, D. (2024, December). Intelligent solar energy harvesting and management in IoT nodes using deep self-organizing maps. In *2024 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1-6). IEEE.
77. Praveen, R. V. S., Hemavathi, U., Sathya, R., Siddiq, A. A., Sanjay, M. G., & Gowdish, S. (2024, October). AI Powered Plant Identification and Plant Disease Classification System. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1610-1616). IEEE.

78. Dhivya, R., Sagili, S. R., Praveen, R. V. S., VamsiLala, P. N. V., Sangeetha, A., & Suchithra, B. (2024, December). Predictive Modelling of Osteoporosis using Machine Learning Algorithms. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 997-1002). IEEE.
79. Kemmannu, P. K., Praveen, R. V. S., Saravanan, B., Amshavalli, M., & Banupriya, V. (2024, December). Enhancing Sustainable Agriculture Through Smart Architecture: An Adaptive Neuro-Fuzzy Inference System with XGBoost Model. In *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)* (pp. 724-730). IEEE.
80. Praveen, R. V. S. (2024). *Data Engineering for Modern Applications*. Addition Publishing House.