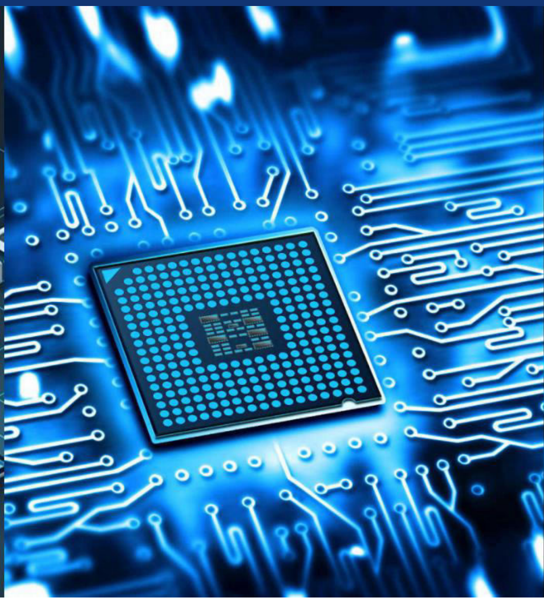
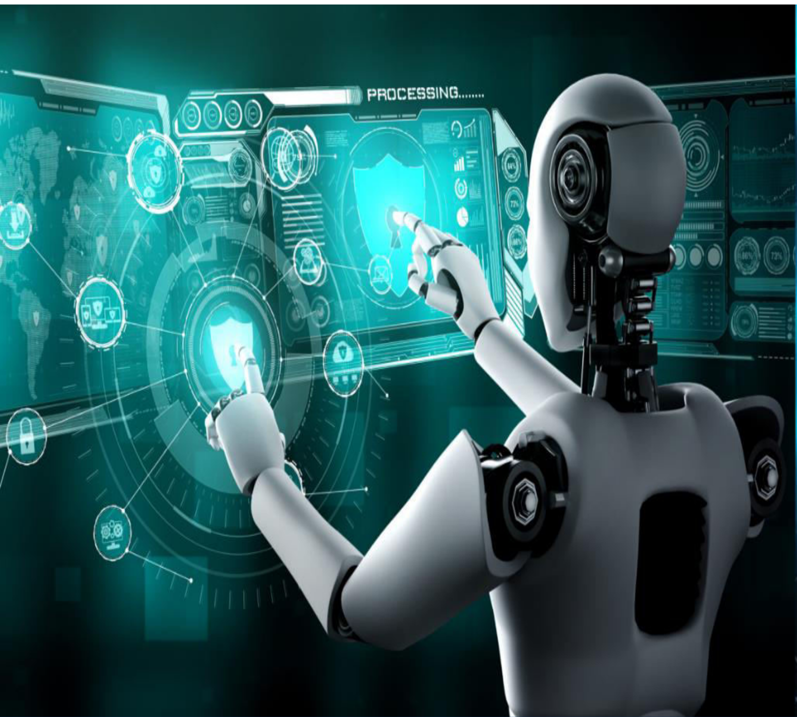
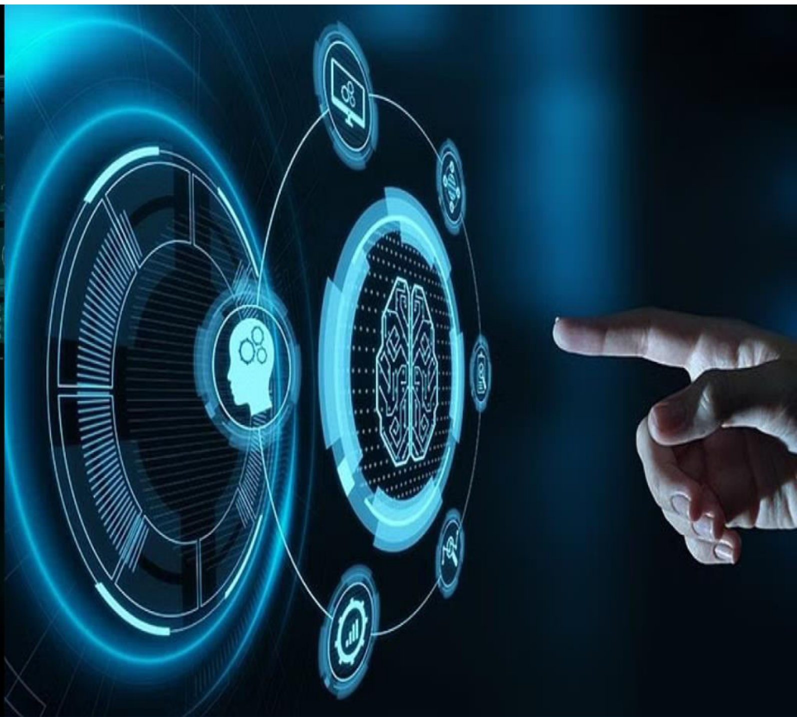
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Enhanced Stock Market Prediction: Ensemble Learning Methods

Shaik Tajuddin[1], Sakhamuri Sushma[2], Pathan Sai Baji[3], Kokkiligadda Sri Chakresh[4], Ch. Sai Siva Durga[5]

U G Students, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[1-4]

Asst. Professor, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[5]

ABSTRACT: Predicting stock prices is difficult because financial markets are complex and change quickly. This project focuses on using advanced machine learning techniques to predict stock price trends. It uses real-time data from Yahoo Finance and the Alpha Vantage API. The goal is to build a smart model that studies past stock prices and other important factors to predict future movements. The model uses machine learning to find patterns in the data, including technical indicators (like moving averages) and broader economic factors, to help investors make better decisions. Both supervised and unsupervised machine learning methods are used to handle large amounts of data. The model updates its predictions regularly with live market data. A user-friendly interface lets investors see the predictions and explore possible investment options.The system also includes sentiment analysis from news and social media to improve accuracy. By combining many data sources, this model gives useful insights to help investors manage their portfolios and reduce risk.

KEYWORDS: Predictive Analytics, Financial Data, Yahoo Finance, Alpha Vantage API, Time Series, Volatility.

I. INTRODUCTION

The stock market is a crucial part of the global financial system, facilitating investment, economic expansion, and capital formation. However, forecasting stock prices is a complex task due to the market's inherent volatility, vast data dimensions, and the influence of external factors such as economic shifts, political events, and social trends. Traditional statistical models often struggle to capture these complexities, leading to an increased reliance on machine learning (ML) techniques for stock price prediction. Machine learning models can process large datasets, identify intricate patterns, and improve forecasting accuracy. This project focuses on developing an intelligent stock price prediction system by integrating historical stock data with ML models, including ARIMA, LSTM, and Linear Regression. Additionally, sentiment analysis on financial news and social media is incorporated to gauge investor sentiment, which significantly impacts stock price movements.

Predictive analytics plays a critical role in financial markets by leveraging historical data, statistical algorithms, and ML techniques to detect trends and make informed predictions. The process begins with data collection, where stock prices, trading volumes, and financial news are gathered from reliable sources. High-quality data is essential for making accurate predictions. The next step involves data processing, where raw data is cleaned, structured, and transformed to eliminate inconsistencies, missing values, and noise. This ensures that models receive well-prepared input, enhancing their ability to generate meaningful predictions.

Feature engineering is an essential aspect of this forecasting process, where key indicators such as moving averages, relative strength index (RSI), and moving average convergence divergence (MACD) are extracted. These technical indicators provide valuable insights into market trends and momentum, serving as input features for predictive models. Following feature extraction, model training begins, implementing different ML algorithms to optimize forecasting accuracy. The ARIMA model is widely used for capturing linear trends and seasonal patterns in time-series data, but it falls short when dealing with nonlinear market behaviours. To address this limitation, deep learning models such as LSTM, a variant of recurrent neural networks (RNNs), are employed. LSTMs are capable of learning long-term dependencies in sequential data, making them highly effective for stock price prediction. Additionally, Linear Regression is utilized as a baseline model to provide a straightforward and interpretable approach to price forecasting.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Once models are trained, their performance is assessed using evaluation metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These metrics help determine which model delivers the highest accuracy. The best-performing model is then deployed for real-time stock market predictions. To further refine forecasting accuracy, sentiment analysis is incorporated. Investor sentiment, derived from financial news and social media discussions, plays a significant role in market behaviour. By using natural language processing (NLP) techniques, sentiment scores are extracted and integrated into prediction models, offering a more comprehensive market outlook.

In the paper, we explained the related works in section-II , Background of which algorithms used and implemented in our ESMP(ENSEMBLE METHODS FOR STOCK MARKET PREDICTION) Model is explained in section-III in detail. Our Proposed System section -IV, Comparative analysis in section-V, Results of proposed Model and Conclusion in section-VI

II. RELATED WORKS

"A Deep Fusion Model for Stock Market Prediction with News Headlines and Time Series Data" by Pengfei Chen, Zisis Boukouvalas, and Riccardo Corizzo (2024): This study introduces a deep fusion model that combines news headlines and time series data to predict stock market trends. The model integrates textual information from news with historical stock data, enhancing prediction accuracy.

"FinBERT and LSTM-Based Novel Model for Stock Price Prediction Using Technical Indicators and Financial News" by Gourav Bathla and Sunil Gupta (2024): This study utilizes a hybrid model combining FinBERT, specifically trained on financial corpora for sentiment analysis, and Long Short-Term Memory (LSTM) networks. The model merges sentiment scores from financial news with technical indicators to predict stock prices, demonstrating improved accuracy during periods of high market fluctuation.

"Enhancing Stock Market Forecasting: A Hybrid Model for Accurate Prediction of S&P 500 and CSI 300 Future Prices" by authors not specified (2025): This paper investigates the development of predictive models to forecast stock prices accurately, focusing on the S&P 500 and CSI 300 indices. It highlights the superiority of ensemble or hybrid models in enhancing prediction reliability, utilizing advanced machine learning techniques and time series analysis

S.no	Paper info	Description	Limitations
1	S. K. Sahoo, S. K. Dash, and S. K. Behera (2023)	This study conducts an exhaustive review of over a hundred research articles focused on global indices and stock prices, analysing various techniques for financial market prediction.	Does not include alternative data sources such as social media sentiment, economic indicators, or news analytics
2	J. D. Hamilton (2023)	This paper reviews current literature on machine learning techniques applied to stock market prediction, aiming to identify directions for future research in this domain.	Potentially overlooking recent advancements in deep learning and hybrid approaches Furthermore, the research lacks a detailed discussion on the practical implementation challenges of these models in real-world trading environments.
3	G. Patel and R. Yalamanchi (2022)	This literature review investigates various machine learning techniques applied to stock market prediction, discussing their effectiveness and the types of data utilized.	Does not explore specific methodologies for improving Stock market prediction accuracy.
4	A. Tesma (2021)	This paper focuses on the application of Recurrent	Recurrent Neural Networks (RNN) and Long Short-Term Memory

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

		Neural Networks (RNN) and Long Short-Term Memory (LSTM) technologies in predicting stock market trends.	(LSTM) models for stock market forecasting. Limited Scope of Data Sources
5	L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan (2023)	This study reviews 30 studies regarding machine learning approaches/models in stock market prediction, highlighting the techniques.	Offers a systematic review of 30 studies on machine learning models used for stock market forecasting Limits to Evaluation Metrics and Benchmarking
6	R. K. Gupta and A. K. Dey (2023)	This paper reviews various methods used to predict stock prices, including technical analysis, time series, and fundamental analysis, discussing their effectiveness and challenges.	Limited Coverage of Machine Learning Techniques
7	X. Li and Y. Wang (2023):	This study provides a comprehensive analysis of several approaches used in stock price forecasting, including statistical, machine learning, and deep learning models, discussing their advantages and limitations.	Practical Implementation Challenges
8	G. Patel and R. Yalamanchi (2022)	This systematic review investigates machine learning techniques applied to stock market prediction, focusing on the markets and stock indices covered, as well as the types of variables used as input.	The research lacks a detailed discussion on the practical implementation challenges of these models in real-world trading environments.
9	Zou, Q. Zhao, Y. Jiao, H. Cao, Y. Liu, Q. Yan, E. Abbasnejad, L. Liu, and J. Q. Shi (2022)	This paper presents a structured and comprehensive overview of research on stock market prediction using deep learning models, proposing a novel taxonomy to summarize state-of-the-art models.	Limits to the Evaluation Metrics and Benchmarking

III. BACKGROUND

1.Machine Learning Models:
1.1 Autoregressive Integrated Moving Average (ARIMA) in Machine Learning

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

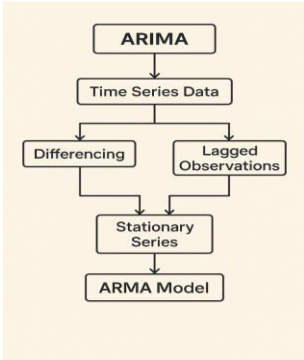


Fig1:ARIMA MODEL

In Fig1 Explain about the ARIMA MODEL The Autoregressive Integrated Moving Average (ARIMA) model is a widely used statistical approach for time series forecasting. It is particularly effective in capturing linear dependencies in sequential data. ARIMA belongs to the class of autoregressive models and is often employed in machine learning applications that involve trend analysis, demand forecasting, and financial predictions.

1.2 LSTM(LONG-SHORT TERM MEMORY):

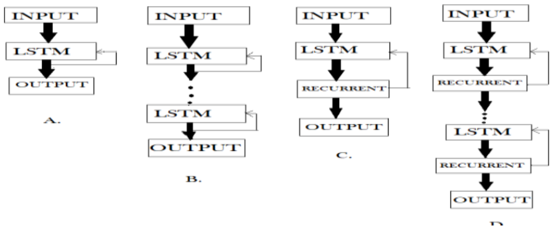


Fig 2.LSTM MODEL

The Fig2 illustrate the model of the LSTM is a type of Recurrent Neural Network (RNN) designed to handle sequential data and address the vanishing gradient problem of traditional RNNs. It is widely used in time-series forecasting, natural language processing (NLP).

Traditional RNNs struggle to learn long-term dependencies because of the vanishing gradient problem. LSTMs solve this by using a memory cell that can store information for long durations and Thus Unlike traditional RNNs, LSTMs overcome the vanishing gradient problem using a memory cell and gates that regulate information flow.

1.3 Linear Regression:

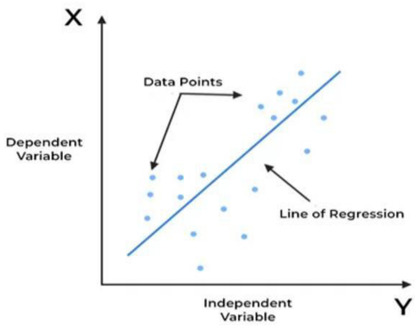


Fig 3 Linear Regression

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304139

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The Fig 3 illustrate about the model Linear Regression is a supervised learning algorithm used for predicting continuous values. It establishes a relationship between the dependent variable (target) and one or more independent variables (features) using a linear equation.

Dataset:

Date	Open	High	Low	Close	Adj_close	Volume	Company
2023-01-18 00:00:00-05:00	97.250000	99.320000	95.379997	95.459999	95.459999	79570400	MRF
2023-01-19 00:00:00-05:00	94.739998	95.440002	92.860001	93.680000	93.680000	69002700	APPLE
2023-01-20 00:00:00-05:00	93.860001	97.349998	93.199997	97.250000	97.250000	67307100	microsoft
2023-01-23 00:00:00-05:00	97.559998	97.779999	95.860001	97.519997	97.519997	76501100	AMAZON
2023-01-24 00:00:00-05:00	96.930000	98.089996	96.000000	96.320000	96.320000	66929500	AMAZON
2023-01-25 00:00:00-05:00	92.559998	97.239998	91.519997	97.180000	97.180000	94261600	Flipkart
2023-01-26 00:00:00-05:00	98.239998	99.489998	96.919998	99.220001	99.220001	68523600	AMAZON

Table 1 .Dataset Format

The above figure explains about the attributes we have taken in the data to efficiently capture the all metrics from the data the attributes are open, high, low ,close, adj_close, volume, company the open attribute captures the price of the stock when the market is opened and the high value captures the stock price up to which the stock price has gone to high and low price captures the lowest price till market is closed and the volume is the number of stock sold and the company name

IV. PROPOSED SYSTEM

Architecture

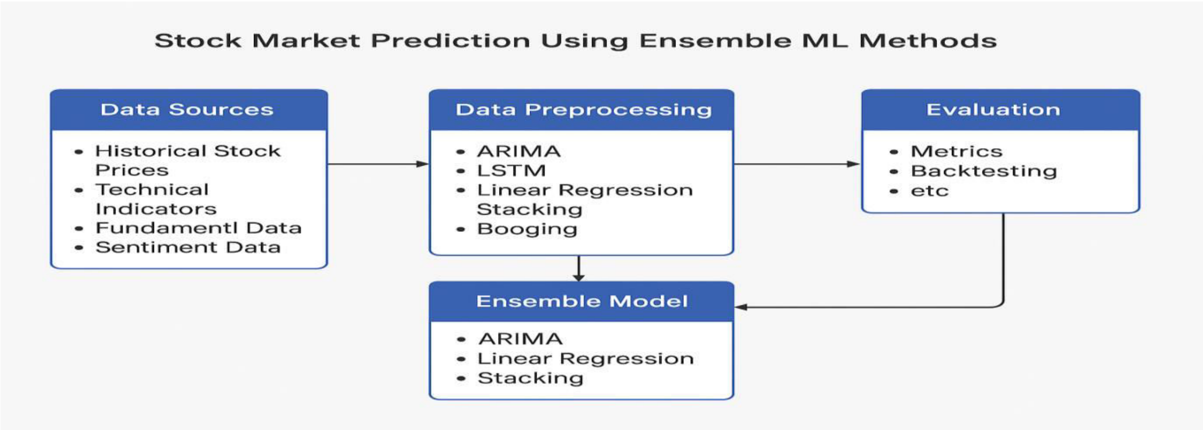


Fig 4 ESMP MODEL ARCHITECTURE

The fig 4 illustrate the architecture of the ESMP MODEL Algorithm outlying the flow of the data and processes involved in identifying the ESMP Model used to predicting the stock pric6e.The system begins with the collection of Stock Data, encompassing various attributes such as open, close and usage patterns, and others .This data serves as the foundation for the predictive model.

The proposed stock market prediction architecture integrates ARIMA, LSTM, and Linear Regression to effectively model both linear and non-linear stock price trends. The process begins with data collection and preprocessing, where historical stock prices—including open, high, low, close values, and trading volume—are gathered and refined to remove inconsistencies. ARIMA is used to detect linear trends and seasonal patterns in the data by analyzing time series characteristics. Meanwhile, LSTM, a deep learning technique, is employed to recognize long-term dependencies and intricate patterns in sequential stock data, improving predictive performance. Additionally, Linear Regression is implemented to establish relationships between stock prices and selected features, serving as a straightforward predictive model.

Each model's performance is assessed using evaluation metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). To enhance prediction accuracy, an ensemble method is applied, combining outputs from ARIMA, LSTM, and Linear Regression through weighted averaging or stacking techniques. The final predictions are then visualized alongside actual stock prices to evaluate the effectiveness of the approach. By leveraging the strengths of these models, this architecture aims to provide a more precise and dependable stock market forecasting system, making it suitable for financial analysis and decision-making.

Workflow:

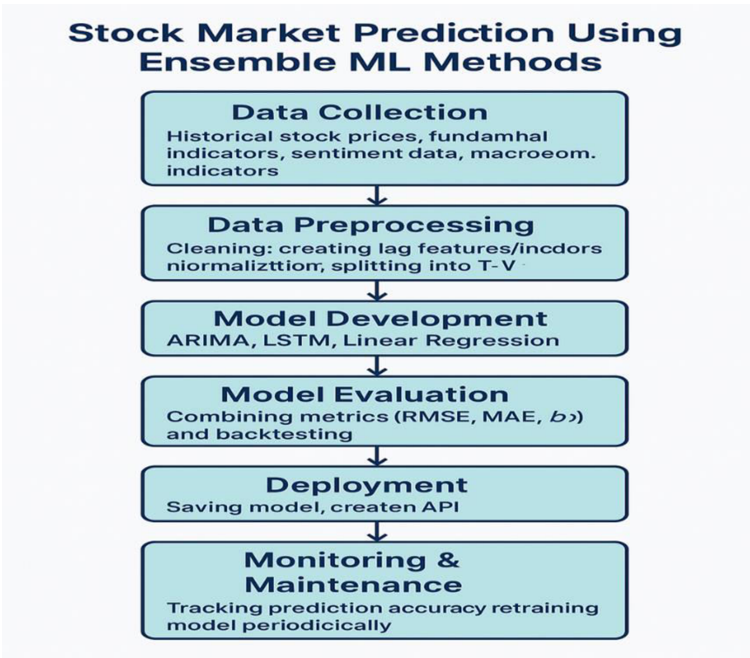


Fig 5 Workflow of ESMP MODEL

The fig 5 illustrate the workflow of your stock market prediction project follows a systematic process, ensuring accurate forecasting through data preparation, model implementation, evaluation, and visualization

1.Data Collection: The process starts by gathering historical stock market data, including open, high, low, and close prices, along with trading volume. If necessary, additional external factors such as economic trends or financial trends are considered Online Data is collected by Using The Alpha_vintage as API
2.Preprocessing: Removing the Null Values and Handling the data as we required for the processing

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.Feature Selection: To improve prediction accuracy, essential features like moving and trend indicators are extracted.
4.Model Selection: Selecting the required models for your project is an essential step for the project considering all the requirements and selecting the model is an crucial part
5.Model Training: Training the Selected with correct Weights or parameters defines the accuracy and how far our proposed tech is good for the future techniques
6.Prediction and Analysing the result: Testing The trained with the test data and predicting the values of the stocks and analysing the error and bias by using the metrics like RMSE and many others

V. COMPARATIVE ANALYSIS WITH OTHER MACHINE LEARNING MODELS IN THE MARKET

Model	**Study**	**Performance Metrics**	**Key Findings**
Ensemble Model (ARIMA, LSTM, Linear Regression)	Our Model ESMP	RMSE:.60.89	Combines ARIMA's capability to model linear trends, LSTM's strength in capturing long-term dependencies, and Linear Regression's simplicity in modelling variable relationships
Random Forest	Brahmanapalli et al. (2024)	MAE: 15.98, RMSE: 27.34	Performs well with structured data but may not effectively capture temporal dependencies, making it less suitable for time series forecasting without modifications.
XGBoost	Aiyegbeni & Li (2024)	R^2: 20.47%, MAE: 15.98, RMSE: 27.34	Demonstrates low predictive accuracy with no proper hyperparameter tuning. No Effective in handling structured data and capturing complex patterns
K-Nearest Neighbours (KNN)	Li (2024)	MAE:20.5 RMSE:30.07	Performs adequately but may lack the capacity to capture complex temporal dependencies in stock price data, limiting its effectiveness in time series forecasting.

Table-2 Comparing ESMP MODEL With other

The table presents a comparative analysis of machine learning models, ranked by test accuracy, showcasing their performance across test precision, ESMP MODEL leads with 60.89% accuracy and consistently high scores in all metrics, indicating robust performance. demonstrating strong but slightly varied results.A performance drop is observed with Random Forest, exhibiting significantly lower accuracy and metrics. XGBOOST and KNN further this trend, with accuracies below 50% Suggesting potential Model challenges.

In summary, CCP Boost Classifier, XGBoost, RandomForest, and KNN are the top performers, with ESMP Model being the most effective. The remaining models indicate a need for further optimization or alternative model selection for improved results.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. RESULT

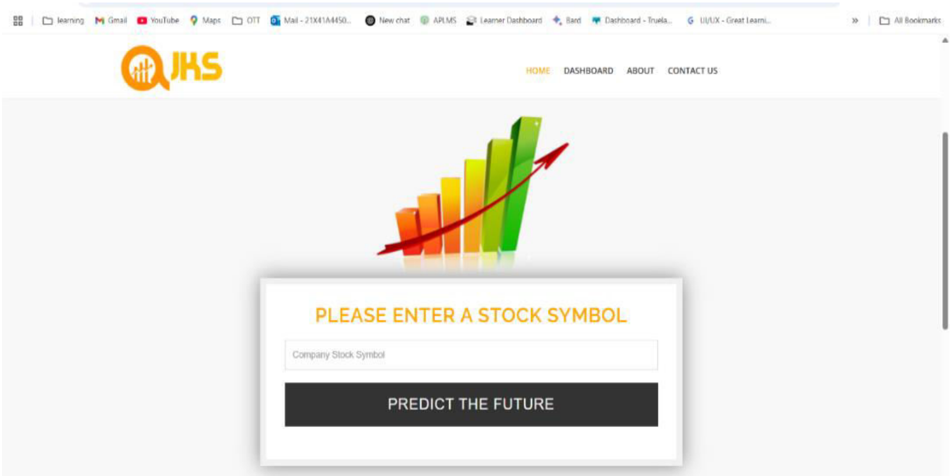
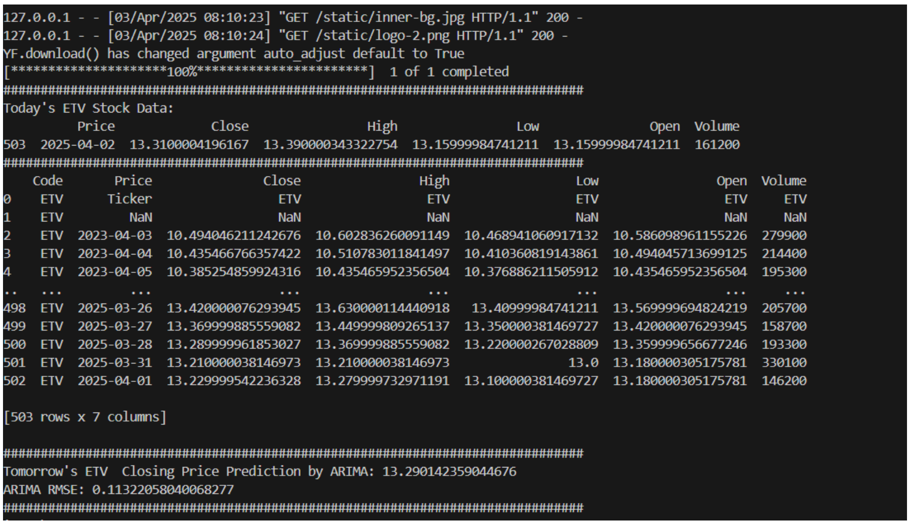


Fig: Input page

The Fig is an input page in which we will enter the stock ticker symbol we required to predict and use our ESMP Model for the prediction



```
127.0.0.1 - - [03/Apr/2025 08:10:23] "GET /static/inner-bg.jpg HTTP/1.1" 200 -
127.0.0.1 - - [03/Apr/2025 08:10:24] "GET /static/logo-2.png HTTP/1.1" 200 -
YF.download() has changed argument auto_adjust default to True
[*********************100%***********************]  1 of 1 completed
##############################################################################
Today's ETV Stock Data:
           Price            Close               High                Low               Open  Volume
503  2025-04-02  13.3100004196167  13.390000343322754  13.15999984741211  13.15999984741211  161200
##############################################################################
    Code       Price               Close                High                 Low                Open  Volume
0    ETV      Ticker                 ETV                 ETV                 ETV                 ETV     ETV
1    ETV         NaN                 NaN                 NaN                 NaN                 NaN     NaN
2    ETV  2023-04-03  10.494046211242676  10.602836260091149  10.468941060917132  10.586098961155226  279900
3    ETV  2023-04-04  10.435466766357422  10.510783011841497  10.410360819143861  10.494045713699125  214400
4    ETV  2023-04-05  10.385254859924316  10.435465952356504  10.376886211505912  10.435465952356504  195300
..   ...         ...                 ...                 ...                 ...                 ...     ...
498  ETV  2025-03-26  13.420000076293945  13.630000114440918   13.40999984741211  13.569999694824219  205700
499  ETV  2025-03-27  13.369999885559082  13.449999809265137  13.350000381469727  13.420000076293945  158700
500  ETV  2025-03-28  13.289999961853027  13.369999885559082  13.220000267028809  13.359999656677246  193300
501  ETV  2025-03-31  13.210000038146973  13.210000038146973                13.0  13.180000305175781  330100
502  ETV  2025-04-01  13.229999542236328  13.279999732971191  13.100000381469727  13.180000305175781  146200

[503 rows x 7 columns]

##############################################################################
Tomorrow's ETV  Closing Price Prediction by ARIMA: 13.290142359044676
ARIMA RMSE: 0.11322058040068277
##############################################################################
```

Fig: Data retrieved

The Fig shows the Retrieved data when we enter the stock ticker symbol in the input page It takes the parameters as open ,high ,low ,close ,volume as the parameters for the entered Stock and the evaluated the ESMP Model

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

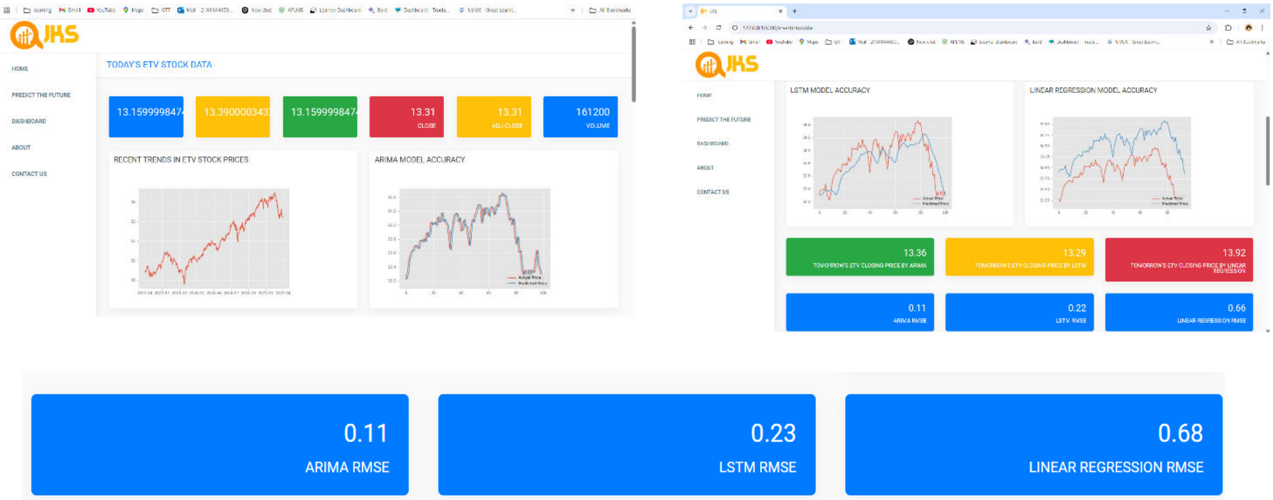


Fig: prediction Page

The fig illustrates the prediction for the stock you entered in the input model it develops an individual visualization for each algorithm used in the ESMP Model and we can Take Decision by understanding this page

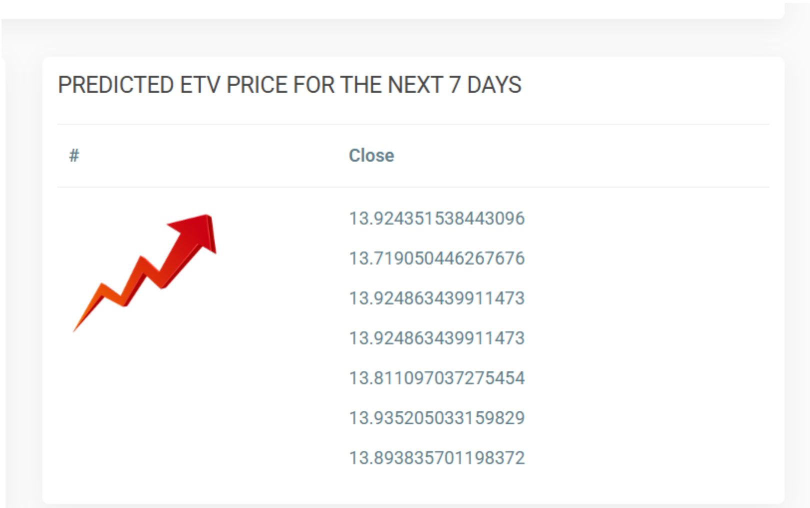


Fig: Next 7 days Stock price

The Fig gives the next 7days price prediction for the stock you entered in the input and we can decided whether we want to buy or sell the stock

VII. CONCLUSION

The stock market plays a big role in the economy, and predicting prices accurately helps investors and traders make better decisions. This project shows how machine learning models like ARIMA, LSTM, and Linear Regression can predict stock prices more accurately by analyzing past data. It also uses sentiment analysis from news and social media to understand market mood. By combining different models and indicators like RSI and MACD, the system gives better predictions and supports smart investing. The project can grow even more by adding new economic data, improving deep learning models, and enhancing sentiment analysis. Overall, it offers a smart, data-driven way to improve investment strategies and reduce risks

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1.”Application of Ensemble Machine Learning in the Predictive Data Analytics of Indian Stock Market” Volume 16, Number 2, December, 2019 Home Table of Contents Titles & Subject Index Authors In

2."Stock Price Prediction Based on ARIMA-BP Neural Network" Published: April 2025

3."Stock-Price-Prediction-Using LSTM ARIMA" Published: January 2025

4."Stock Price Prediction and Comparative Analysis Using RNN, LSTM, and ARIMA Models" Published: January 2025

5."Prediction of Stock Market Using Sentiment Analysis and Ensemble Techniques" Published: March 2025

6."Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets" Published: April 2022

7."A Hybrid Model Integrating Deep Learning with Investor Sentiment Analysis for Stock Price Prediction Published: 2021

8."Using Deep Learning Techniques in Forecasting Stock Markets by Hybrid Data with Multilingual Sentiment Analysis" Published: 2022

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING