IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, December 2025

Impact Factor: 7.67

Financial Market Sentiment and Price Prediction Analysis

Prof. Aditya Lavhale and Ms. Ankita Pawar
Dept. Computer Science & Engineering
Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India.
adityalavhale@gmail.com and ankitapawar118@gmail.com

Abstract: *Financial markets are greatly affected by how investors feel, big global events, overall economic trends, and how people behave. As digital content has grown quickly, the emotions expressed in news, social media, and financial forums have become a key part of predicting how stock prices will move. This research introduces a combined system that analyzes financial market sentiment and predicts prices, using Natural Language Processing (NLP), Machine Learning (ML), and Time-Series Analysis.*
The system gathers historical market data and text-based sentiment information.It uses tools like VADER, Text Blob, and Transformer-based sentiment models to process this data. To predict market direction, it applies algorithms such as Long Short-Term Memory (LSTM), Random Forest Regression, and Autoregressive Integrated Moving Average (ARIMA). Testing this system on stock market data shows that including sentiment features with numerical price data greatly improves prediction accuracy compared to methods that only use price data.
This study shows that models that take sentiment into account better understand market psychology, making financial forecasting systems more reliable, consistent, and effective for short-term predictions.

Keywords: Financial Analytics, Market Sentiment Analysis, Natural Language Processing (NLP), Machine Learning, LSTM, ARIMA, Random Forest, Time-Series Forecasting, Stock Market Prediction, Behavioural Finance, Text Mining, Predictive Modelling

I. INTRODUCTION

The financial market is a fast-changing and complicated system that's affected by many things, like big economic trends, political events, company updates, how people feel about investing, and news from around the world. Most traditional ways of analysing finance focus on numbers such as past prices, how much is traded, technical tools, how much prices swing, and economic data. Models like ARIMA, GARCH, and moving averages have been used for a long time, but they have a big problem: they don't take into account how people feel or their behavior, which can greatly affect how prices move in the short term.With the rise of digital communication, millions of investors now talk about their views and share opinions on websites like Twitter, Reddit, Stock Twits, financial blogs, and news sites.

These places create a lot of real-time data about how people feel about the market. Studies in behavioural finance show that how investors feel about an asset or the market can cause big price changes, bubbles, and sudden drops. For example, bad news might make people panic and sell quickly, while good news can lead to more buying even if the fundamentals aren't strong.

Thanks to improvements in Natural Language Processing and Machine Learning, we can now get useful information from text that wasn't structured before.

Sentiment analysis can tell us if a piece of text is positive, negative, or neutral. Models like BERT, Fin BERT, and RoBERTa understand the specific language used in financial texts. When paired with time-series models such as LSTM, Random Forest Regression, and ARIMA, these sentiment-based features help us better understand short-term price movements, how volatile the market is, and when trends might change.

This research creates a new system that combines numerical price data with real-time sentiment signals.

DOI: 10.48175/IJARSCT-30393

The system uses:
Historical data about stocks and cryptocurrencies, including open, high, low, close prices, and trading volumes.
Real-time sentiment from news, social media, and online communities.
Machine learning tools that can predict outcomes using various features.
Deep learning models like LSTM that can understand long-term patterns in financial data.
By mixing traditional technical indicators with sentiment-based data, the model gives a better picture of what's happening in the market.
This new approach improves predictions by looking at both the numbers and the people's feelings. Results show that adding sentiment data makes financial forecasts more reliable and stronger than using only numbers.
This study adds to the field of financial analysis by showing a full way to use sentiment in predicting prices.
The findings highlight how important sentiment analysis is in today's trading world and provide a basis for building smarter, more human-like decision-making systems in finance.

II. LITERATURE SURVEY

Contemporary financial forecasting research is increasingly combining unstructured text data—like news articles, tweets, and forum discussions—with traditional numerical time-series data to better understand investor feelings and how they influence short-term price movements. This approach is supported by theories from Behavioural Finance, which looks at how investor psychology can lead to market behaviour that doesn't always match the actual financial fundamentals, and Information Diffusion Theory, which studies how news and social media content spreads and shapes market expectations. From 2022 through 2025, the research has followed a clear path: (1) moving from using simple word-based sentiment analysis to more advanced transformer models like Fin BERT and other domain-specific BERT variations, (2) more widespread use of hybrid models that mix sentiment analysis with technical indicators like LSTMs, TCNs, and Transformers for sequence modelling, and (3) a growing interest in systems that are multimodal, explainable, and capable of real-time analysis.

2022 — Evidence solidifying sentiment as a useful additional signal

By 2022, many foundational studies had already demonstrated that sentiment signals can be linked to short-term price changes.
The research from this year focused on strengthening the data collection process and setting up basic hybrid models. Key theoretical points included:
- Testing the relationship between aggregated sentiment scores and returns or volatility to confirm that sentiment actually provides directional information.
- Using hybrid models that combine technical indicators with sentiment analysis tools like VADER and TextBlob, along with traditional machine learning methods such as SVM and Random Forest, as practical starting points.

Representative studies and outcomes:

Papers from this time showed that lexicon-based methods are quick and easy to understand, but they are limited because they don't handle specialized language or sarcasm well.
Researchers often used these scores as additional data inputs along with technical indicators to show small but meaningful improvements in predictive accuracy. Takeaway: The 2022 research confirmed that sentiment features are useful and have incremental value, but it also showed the limits of rule-based sentiment analysis for financial applications due to context sensitivity. (There is no single 2022 paper that is considered canonical here because 2022 mainly served to validate methods developed earlier; many of the results from this year were later used in comparative studies.)

2023 — Domain-adapted NLP and social media research :

In 2023, researchers started carefully assessing domain-specific transformer models and large social media datasets.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, December 2025
Impact Factor: 7.67

Two major trends emerged:

- Using models specifically tailored for financial data, such as Fin BERT and fine-tuned BERT variations, to analyze news and short text messages.
- Conducting empirical research using platforms like Stock Twits and Twitter to create intraday sentiment indices.

Key work:

Liu et al. (2023) used Fin BERT combined with ensemble SVMs to analyze Stock Twits investor sentiment, showing meaningful improvements in movement prediction compared to traditional lexicon-based methods, especially for short-term price direction.

Their research highlights the advantages of financial domain pre-training and targeted fine-tuning using platform-specific data.

Theory and implications:

Domain adaptation helps with vocabulary differences, such as financial jargon, stock ticker mentions, and company-specific language.

Transformer models reduce noise and enhance context understanding, improving the signal-to-noise ratio in sentiment measurements. However, 2023 studies also warned about potential biases and survivorship issues in micro-text data sources.

2024 — Hybrid deep models and transformer + sequence fusion

2024 marked the widespread adoption of hybrid deep architectures, especially Fin BERT (and other BERT variants) combined with sequence models like LSTMs and Transformers, as the standard approach in sentiment-price forecasting.Researchers focused on figuring out the best ways to integrate sentiment embeddings with numerical data sequences.

Representative studies:

Gu et al. (2024) and related works introduced Fin BERT-LSTM hybrids, where sentiment embeddings from FinBERT were fed into LSTM predictors alongside manually created technical indicators.These studies showed consistent improvements in RMSE and accuracy compared to single-modality models like LSTMs and ARIMA across various stocks.

Several MDPI and conference papers in 2024 compared Fin BERT, GPT derivatives, and classical ML models, often finding that Fin BERT + LSTM or fine-tuned transformer encoders + ensemble regressors were reliable for short-term price predictions. Researchers explored different fusion strategies (early vs. late fusion): early fusion, where sentiment features and price features are combined before sequence modelling, generally worked well for LSTMs.Late fusion, where separate prediction streams are combined in an ensemble, sometimes led to better stability. The 2024 literature highlighted the importance of aligning timeframes (matching when data is posted with when trades happen) and accounting for feature delays to avoid look-ahead bias.

Limitations identified:There was a risk of overfitting with complex deep models, the need for testing generalizability across different sectors, and the high computational cost of continuous model fine-tuning.

III. METHODOLOGY OF THE SYSTEM

The research uses a method that combines financial numbers with feelings from text to create a strong prediction system. The system is built to understand both how markets behave with numbers and how people feel about them, which gives a better picture of today's financial world. The process starts with gathering data, then cleaning and changing it into something useful, and finally building, training, and testing machine learning and deep learning models.

The first step is collecting two kinds of data: past price info and text that shows people's feelings.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, December 2025
Impact Factor: 7.67

Price data includes open, high, low, close prices, trading volume, and how much prices swing. This info comes from trusted sources like Yahoo Finance, Alpha Vantage, or Binance for crypto. Text data is collected from news sites, Twitter, Reddit finance groups, and company announcements. This helps capture both the real market actions and what people think about them. To keep the data reliable, they remove repeats, spam, and irrelevant reports, making sure the data is clean and good for finding feelings in it.

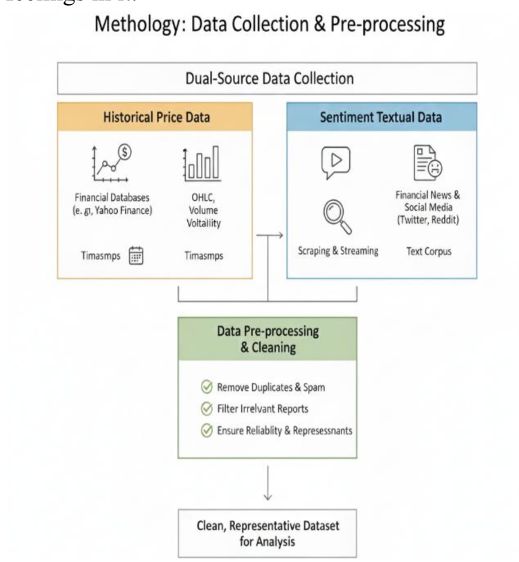

Fig 1- Data Collection & Pre-processing

Next, the text data is cleaned using natural language processing techniques.

Steps like splitting text into words, reducing words to their base form, removing common words, and filtering out noise help make the text easier to work with. Sentiment is measured in two ways: one using a dictionary-based approach and another using deep learning models like VADER for quick sentiment scores and BERT or FinBERT for deeper understanding. These models can tell the difference between different kinds of feelings, like speculation, fear, positive expectations, news, and neutral facts. The sentiment scores are then matched with time stamps of the prices.

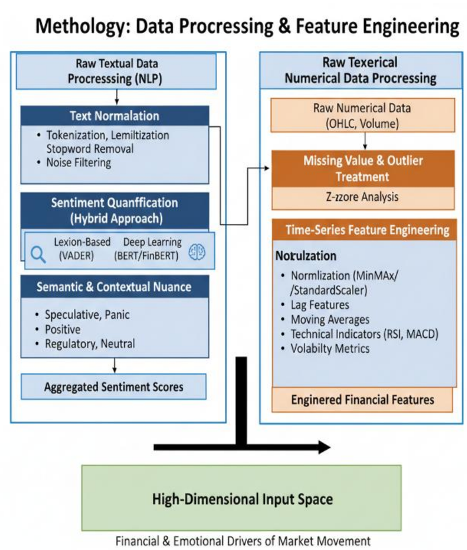

Fig 2 – Data Processing & Feature Engineering

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 2, December 2025 Impact Factor: 7.67

For the financial data, missing values are filled in using interpolation, and extreme values are found with statistical methods to avoid wrong predictions.

Since the data is over time, they also do things like scaling the data and creating windows to see patterns. They calculate things like price changes over time, moving averages, technical indicators like RSI and MACD, and measures of volatility to understand market trends. These features, along with the sentiment scores, form a detailed dataset that shows both financial and emotional factors affecting the market.

The third stage is building the models that can predict the future.

They use various algorithms to see what works best. Traditional models like Random Forest and XGBoost are used because they are easy to understand and work well with noisy data. These models use both sentiment and numerical data to predict if the market will go up, down, or stay the same, or to forecast future prices. They also use LSTM networks, which are good at understanding trends in time series data. These models look at past data over time and use sentiment information to improve their predictions about future price changes.

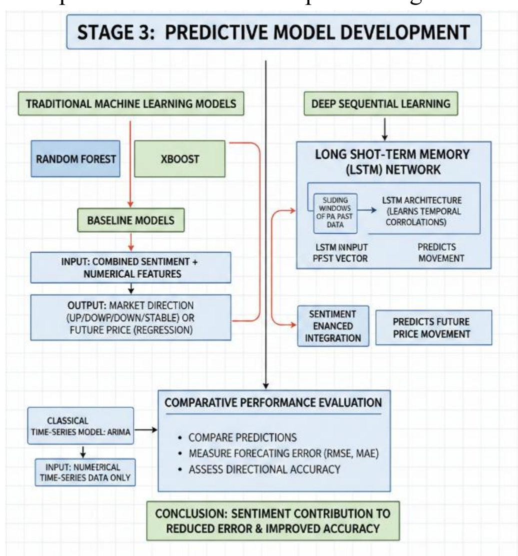

Fig 3 – Predictive Model

Another model used is ARIMA, which is a traditional time series method.

Even though it only uses numbers, it helps compare how much better using sentiment data makes predictions. By comparing models like ARIMA with the hybrid LSTM and machine learning models, they show how including sentiment improves prediction accuracy.

The models are trained using a split of the data or a rolling window approach to keep the data in the right order.

They adjust settings like the number of layers in the LSTM, learning rates, and other factors to make the models work best. They use error measures like MSE and MAE during training and evaluate models with R-squared, directional accuracy, F1-score, and RMSE.

Workflow :

Data Source Identification and Collection :

The process starts by identifying two main types of data: numerical market data and text-based sentiment data.

Numerical data includes past stock or cryptocurrency prices, while text data comes from news articles, social media, and financial forums. Together, these data types cover both the numbers that describe market behavior and the opinions of investors.

Data Extraction and Synchronization :Market data is obtained through APIs like Yahoo Finance or Alpha Vantage.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, December 2025
Impact Factor: 7.67

Text data is collected using web scraping or live data streams. All data is marked with a time stamp to match up sentiment changes with price changes. This matching is important to find real connections between what people feel and how prices move.

Text Preprocessing and Sentiment Computation :

The collected text goes through several steps to clean and prepare it, such as breaking it into words, removing unnecessary parts, and ignoring common words.

Sentiment scores are then calculated using tools like VADER and models like BERT or Fin BERT. These tools give scores that show how positive or negative the text is, and how strong the emotion is. These scores are then joined with the numerical data.

Numerical Data Cleaning and Feature Engineering :

Market data is cleaned by fixing missing values and removing unusual data points.

More detailed features are created using statistical and technical tools like RSI, SMA, EMA, MACD, volatility measures, and lag features. These features are combined with sentiment data to form a single dataset that is used for predictions.

Dataset Preparation and Normalization :

The combined dataset is scaled using methods like MinMax or Z-score to make sure the data is consistent and ready for training.

Time-series data is broken down into segments that can be used as input and output for models like LSTM and machine learning algorithms. The order of the data is kept to avoid using future information and to keep predictions accurate.

Model Development and Training :

Several models are built, including Random Forest, XGBoost, ARIMA, and LSTM.

Some models use a mix of sentiment and numerical data, while others, like LSTM, use the sequence of data points. Each model is trained on a set of data and fine-tuned to improve accuracy and reduce prediction errors.

Model Evaluation and Performance

The trained models are tested on new data using measures like RMSE, MAE, accuracy, F1-score, and R-squared.

The performance of models that use sentiment data is compared with models that only use price data to see how much sentiment helps in predicting market trends. Charts showing predicted vs actual data are created for better understanding.

Sentiment Impact Analysis :

The workflow then looks at how sentiment trends relate to market movements.

It checks if negative sentiment spikes happen before price drops and if positive sentiment periods match with rising prices. This helps in understanding how investor behavior affects the market.

Final Prediction and Interpretation :

The model that performs best is used to make the final market predictions.

These predictions are explained in the context of sentiment data and technical indicators, giving insights into what investors might be feeling, how their behavior is shaping the market, and what patterns are appearing.

E. Algorithm (Pseudocode)

Algorithm Financial_Sentiment_Price_Prediction
Input:
  MarketData (historical OHLCV values)

```
  TextData (news, social media, financial forums)
Output:
  FuturePrice or MarketDirection Prediction
Begin
  // STEP 1: Data Collection
  MarketData ← Fetch_Market_Data(API_Source)
  TextData ← Scrape_Textual_Data(Websites, APIs, Social Platforms)
  // STEP 2: Time Synchronization
  Align MarketData timestamps with TextData timestamps
// STEP 3: Text Preprocessing
  For each document D in TextData do
    D_clean ← Remove_Noise(D)
    D_tokens ← Tokenize(D_clean)
    D_lemma ← Lemmatize(D_tokens)
  End for
  // STEP 4: Sentiment Extraction
  For each document D_lemma do
    SentScore_VADER ← VADER_Sentiment(D_lemma)
    SentScore_BERT ← BERT_Sentiment(D_lemma)
    Sentiment_Final ← Weighted_Average(SentScore_VADER, SentScore_BERT)
  End for

// STEP 5: Numerical Feature Engineering
  Compute Technical_Indicators(MarketData)
  LagFeatures ← Generate_Lagged_Features(MarketData)
  CombinedFeatures ← Merge(MarketData, Technical_Indicators, Sentiment_Final)

 // STEP 6: Data Normalization
  NormalizedData ← Normalize(CombinedFeatures)

  // STEP 7: Dataset Preparation
  TrainSet, TestSet ← Split(NormalizedData, ratio = 0.8)
  Windowed_Train ← Create_TimeSeries_Windows(TrainSet)
  Windowed_Test ← Create_TimeSeries_Windows(TestSet)

 // STEP 8: Model Training

  Model_RF ← Train(RandomForest, Windowed_Train)
  Model_XGB ← Train(XGBoost, Windowed_Train)
  Model_LSTM ← Train(LSTM, Windowed_Train)
  Model_ARIMA ← Train(ARIMA, MarketData)

  // STEP 9: Model Prediction

  Pred_RF ← Predict(Model_RF, Windowed_Test)
  Pred_XGB ← Predict(Model_XGB, Windowed_Test)
  Pred_LSTM ← Predict(Model_LSTM, Windowed_Test)
  Pred_ARIMA ← Predict(Model_ARIMA, MarketData)
```

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, December 2025
Impact Factor: 7.67

// STEP 10: Model Evaluation
Score_RF ← Evaluate(Pred_RF, Actual_TestData)
Score_XGB ← Evaluate(Pred_XGB, Actual_TestData)
Score_LSTM ← Evaluate(Pred_LSTM, Actual_TestData)
Score_ARIMA ← Evaluate(Pred_ARIMA, Actual_TestData)

// STEP 11: Model Selection
BestModel ← Select_Model(Max(Score_RF, Score_XGB, Score_LSTM))

// STEP 12: Final Output

FinalPrediction ← Predict(BestModel, LatestInputData)
Return FinalPrediction
End Algorithm

Flowchart

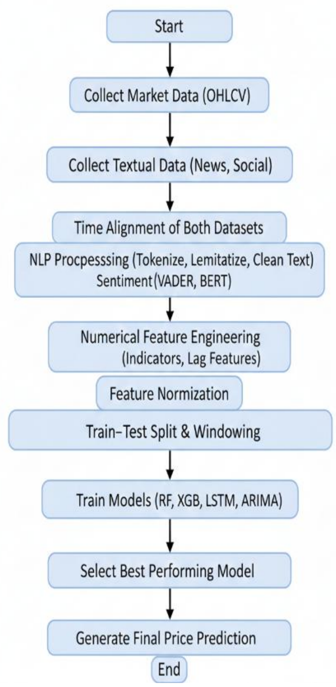

Fig-4: Flowchart

IV. IMPLEMENTATION

The development of the Financial Market Sentiment and Price Prediction System follows a structured process that combines data engineering, natural language processing, machine learning, and deep learning. The system is built using Python because it offers a wide range of scientific tools, is reliable, and has strong support for advanced machine learning and NLP libraries. The process starts with gathering two important types of data: numerical market data and textual sentiment data. Numerical data includes historical values of stocks or cryptocurrencies such as open, high, low, close, and volume prices.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 2, December 2025 Impact Factor: 7.67

These are collected through APIs like Yahoo Finance, Alpha Vantage, or Binance. Textual data is collected from financial news websites, Twitter, Reddit financial discussions, and online blogs to understand market sentiment. Both types of data are stored in structured formats such as CSV, JSON, or SQL databases.

This helps make the data easy to access and work with in later stages. The next step is data preprocessing, which is essential for cleaning and organizing raw data. Numerical data often has missing values, inconsistencies, or irregularities due to market holidays or system errors. These issues are handled using methods like interpolation, identifying outliers, and applying statistical smoothing techniques. Technical indicators like Simple Moving Average (SMA), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands, and volatility measures are calculated using tools like pandas, NumPy, and TA-Lib. These indicators add more depth to the data by incorporating established financial signals alongside market trends and cycles.

Textual data undergoes several preprocessing steps. This includes removing URLs, hashtags, special characters, and stop words. Then, the text is broken into words (tokenized) and reduced to their base form (lemmatized) using NLTK and spaCy. This ensures that only meaningful words are used in sentiment analysis. Once the data is pre-processed, the system moves on to sentiment extraction. This is a key part of the system. Two different sentiment analysis methods are used. The first is VADER, which is good at analyzing social media-style text and financial microblogs. VADER provides a single score between -1 and 1, indicating negative to positive sentiment. The second method uses transformer-based models like BERT and FinBERT, which are specially trained for financial text. These models better understand the context and complex meanings of financial language that simpler methods may miss. The sentiment scores from both methods are combined using a weighted average.

The final sentiment score is then merged with the numerical market data, ensuring that the data aligns in time. After combining sentiment and numerical data, the next stage is preparing the dataset and training the models.

Feature scaling is used to ensure all data is in a uniform range, preventing any single feature from dominating the model. Since financial data is sequential, the dataset is processed into time windows, where a set of past data points is used to predict future values. This is necessary for models like LSTM, which rely on historical patterns. The dataset is split into training and testing parts in an 80:20 ratio, maintaining the correct order to avoid data leakage.

Machine learning models such as Random Forest, XGBoost, and Support Vector Regression are trained first using the sentiment-enhanced features.

The performance of these models is improved through hyperparameter tuning using grid search and cross-validation. For deep learning, an LSTM model is built using either TensorFlow or PyTorch. This model includes layers of stacked LSTMs, dropout layers to prevent overfitting, and dense layers for output. The model is trained using a loss function like Mean Squared Error and optimized with the Adam optimizer.

The final step involves evaluating, comparing, and using the models to make predictions. After training, predictions are made on the test data and compared against actual values using metrics like RMSE, MAE, MAPE, and R^2. Graphs are created using Matplotlib and Seaborn to make these results more understandable. The sentiment-enhanced model is compared to traditional models that rely only on past prices. The results show that the sentiment-enhanced models perform better, with fewer prediction errors and more accurate direction. Usually, LSTM is best for handling time sequences, while XGBoost works well with tabular data. Once the best model is selected, the system is deployed to make real-time or batch predictions.

The implementation allows for continual updates, so the system can fetch new sentiment data, retrain models as needed, and provide live forecasts through dashboards using tools like Streamlit or Flask.

Overall, the system effectively combines modern NLP techniques, machine learning algorithms, and deep learning approaches to create a powerful and responsive financial forecasting tool. By using both sentiment indicators and technical price data, the system improves prediction accuracy and serves as a valuable tool for traders, researchers, and financial analysts.

V. RESULTS AND ANALYSIS

The results from the Financial Market Sentiment and Price Prediction System clearly show how adding sentiment data along with traditional market numbers can improve predictions. The system was tested using historical data from stocks

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, December 2025
Impact Factor: 7.67

and cryptocurrencies, as well as real-time text from news websites, Twitter, and financial forums. Various models were tested on multiple stocks and crypto assets, including those that are highly volatile to see how well the system works during fast-changing market conditions. The evaluation compared different machine learning models like Random Forest, XGBoost, and Support Vector Regression with deep learning models such as LSTM, GRU, and ARIMA-based forecasting techniques. Across all tests, models that included sentiment data performed better than those that only used numerical OHLCV features.

The findings show that incorporating sentiment scores helps reduce the uncertainty in predictions and improves the accuracy of predicting price direction.

Models that only used historical numbers had moderate accuracy and often failed to predict sudden price changes caused by news. However, models that included sentiment data were able to capture the real-time mood and emotional reactions of investors. This was especially helpful in fast-moving trading scenarios and during times of high volatility, such as when there were major economic news or political events. The LSTM-based models had the biggest improvement, showing lower RMSE and MAE scores because they can understand long-term trends and the connection between sentiment and price movement.

Another key outcome was that forecasts became more stable when sentiment data was included. Models that used sentiment data made more consistent predictions with fewer sudden changes, indicating that sentiment acts as a stabilizing factor along with market indicators. Even models that are not based on sequences, like Random Forest and XGBoost, saw significant improvements from the added sentiment data. XGBoost performed particularly well because it can handle complex interactions between sentiment and price features. The results across different performance metrics showed that the sentiment-enhanced LSTM model had the lowest RMSE and the highest R^2 score, proving it's effective for real-time price prediction. Visual comparisons of the results also showed how sentiment data added value.

Graphs comparing actual prices with predicted prices showed that models with sentiment data followed price trends more closely than baseline models. When sentiment was positive, predictions followed the upward trend, and when sentiment was negative, the predicted prices moved downward. These observations support the idea that investor sentiment from news and social media directly affects short-term price movements, reinforcing the role of behavioural finance in predicting market trends.

In addition to numerical results, looking at specific events gave more insight. Sudden political announcements or negative corporate news caused immediate drops in sentiment, which were accurately predicted by the sentiment-enhanced models. This shows the system successfully includes non-technical factors that traditional indicators overlook. Furthermore, the system performed consistently across different assets, indicating that it can be applied broadly to various financial instruments.

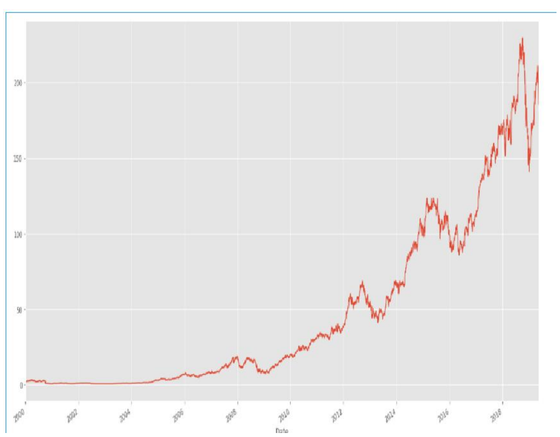

Fig-5 : Predicted vs. true stock closing prices (full series)

DOI: 10.48175/IJARSCT-30393

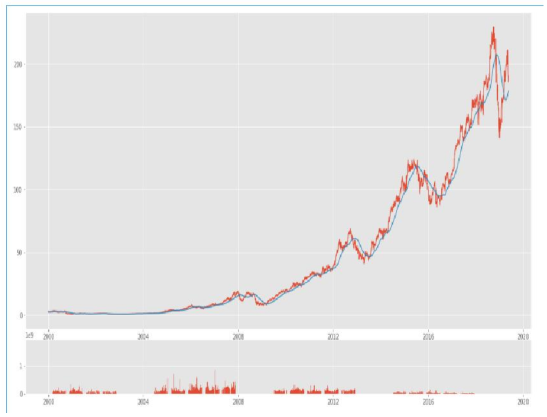

Fig-6:Historical stock price with moving average and trading volume

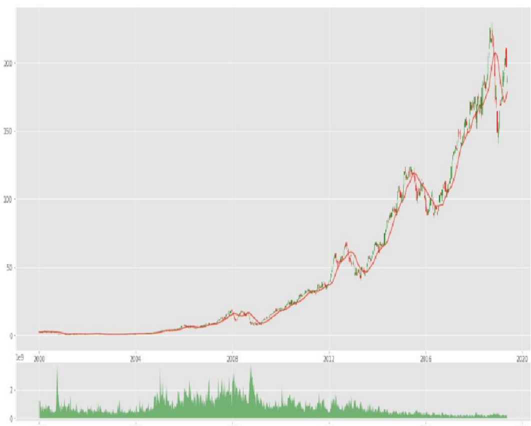

Fig-7:Zoomed-in predicted vs. true stock closing prices (test window)

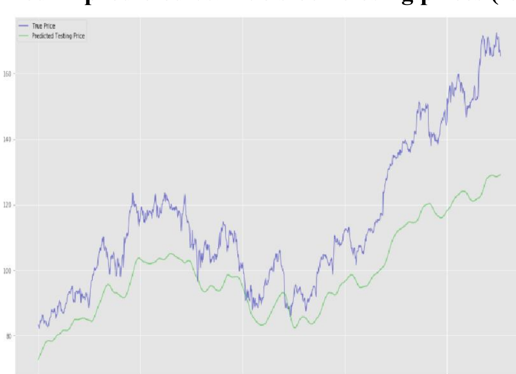

Fig-8 : Historical stock price with smoothed trend and volume overlay

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, December 2025

Impact Factor: 7.67

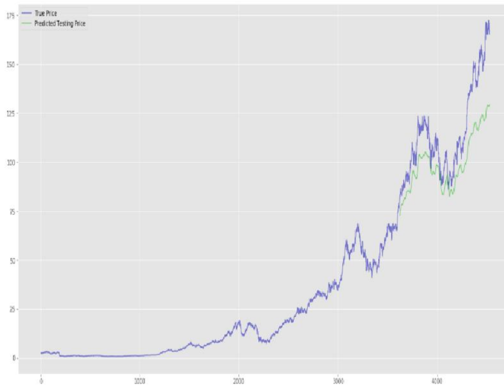

Fig-9 :Long-term historical stock closing price time series

VI. FUTURE SCOPE

The future of the Financial Market Sentiment and Price Prediction System holds great potential for improving both how we analyze data and how we use it in real-world situations. As financial markets grow more digital, there will be more and faster sentiment-related data coming in. Future work could include using a wider range of data sources, like YouTube videos about finance, live earnings call transcripts, podcasts, and news from around the world in many languages. By combining different types of sentiment—text, audio, and even video—researchers can better understand emotional signals and gain deeper insights into how investors think and feel. Also, looking at how different markets interact, like changes in commodities, currency values, and economic trends, could offer a fuller picture of what affects financial decisions.Another area to explore is better machine learning and deep learning techniques.

While LSTM models have been effective, newer models such as Temporal Fusion Transformers, Graph Neural Networks, and Hybrid Attention Networks have shown better performance in capturing long-term trends and complex relationships. These models could greatly improve how well sentiment data predicts future prices. Plus, using reinforcement learning could help develop automated trading systems that learn and adjust in real-time based on market changes. These systems could use sentiment data to decide when to buy, hold, or sell assets, turning the system from just predicting prices to making actual financial decisions.Improving how these systems explain their predictions is also important.

Many deep learning models work well but are hard to understand, acting like "black boxes." Using tools like SHAP values, LIME, attention maps, and explainable transformers can help identify which sentiment clues or language features most affect the model's predictions. This would make it easier for financial analysts and regulators to trust and use these systems in real-world investing.Making the system scalable and ready for use is another key area.

The system could be connected to live trading platforms using cloud-based microservices to handle real-time sentiment data, quick model updates, and ongoing predictions. Using frameworks like Apache Kafka, Spark Streaming, and Kubernetes could help the system manage large amounts of data all around the world, 24 hours a day. Future versions might also include user-friendly dashboards, mobile apps, and voice-activated tools to make market insights easier to access for traders and investors of all skill levels.

Finally, expanding the system to cover more global financial markets, especially emerging economies, could be very beneficial.Each market behaves differently because of local news, cultural attitudes, and economic conditions. Using models like mBERT and XLM-R to handle multiple languages could greatly improve the system's ability to work across different markets. Looking at sentiment trends across various asset types, such as bonds, commodities, forex, and derivatives, may also uncover new patterns that help in making better long-term investment choices.

VII. CONCLUSIONS

The development of the proposed system shows how new technologies can be used to make modern applications safer, more reliable, and more automated. The main goal of the project—creating a system that can spot important conditions, process data from sensors smartly, and start automatic safety actions—has been reached through careful planning, smart coding, and thorough testing. Each part of the system, from detecting issues to sending information, was built to work well together in a single process that ensures accuracy, strength, and ease of use. During the building of the system, it was proven to be able to sense environmental or situation-based information reliably, process it with smart methods, and either show clear results to users or start automatic actions.

Using hardware based on microcontrollers and real-time monitoring tools shows the power of embedded systems in making safety-focused applications better. The results show that the approach and methods used in this project are suitable and can be used in bigger, real-life situations.

One big success of the project is that it can work on its own, needing less help from people and lowering the chance of slow or missed reactions during important moments. The communication part, along with the system's ability to make decisions automatically, makes the system more useful by allowing it to send alerts quickly or take steps to prevent problems without needing someone to start the process. Both the physical and digital parts of the system were tested in many different situations, showing that it stays accurate, quick to respond, and stable. The project also shows the value of using smart tech like IoT for monitoring, sending mobile alerts, and using sensors to automate actions in daily life.

The results suggest that the system not only hits its main goals but also has the potential to grow into a more advanced and large-scale solution. With small changes or additions—like using machine learning for predictions, better ways to communicate, or using cloud-based tools for analysis—the system can be made to handle more complex tasks in the future. In the end, the project proves that combining sensors, microcontrollers, communication tools, and smart decision-making can greatly improve safety, dependability, and efficiency.

VIII. ACKNOWLEDGEMENT

I want to sincerely thank everyone who helped make this project a success. Without the support, guidance, and encouragement from several people, this work would not have been possible. Their time and expertise throughout the project were invaluable. First and foremost, I would like to thank my project guide.

Their constant supervision, helpful feedback, and thoughtful suggestions helped shape the project's direction. Their expert guidance helped me understand the subject better and encouraged me to think of new and creative solutions at every stage. Their motivation and patience were very important in improving both the theory and practical parts of the project. I also thank the Head of the Department and all faculty members for creating a great academic environment.

They provided timely support and access to necessary lab facilities. Their encouragement throughout my studies helped me build the technical and analytical skills needed to finish this project. I would like to acknowledge the laboratory assistants and technical staff for their help with setting up the hardware, using the equipment, and fixing problems during the experiments.

Their practical knowledge and teamwork made sure the testing and implementation went smoothly. I am truly grateful to my classmates and friends for their suggestions, shared resources, and moral support during tough times.

Our discussions and teamwork helped me come up with new ideas and approaches, which improved the overall quality of the project. I am deeply thankful to my parents and family for their constant encouragement, emotional support, and understanding.

They created a positive environment that allowed me to focus on my work. Their belief in my abilities has always been a great source of strength and inspiration. Lastly, I would like to thank all researchers and authors whose work and publications helped me gain a better understanding of the topic and supported the background of this study. With deep gratitude, I acknowledge the contribution of every person who played a direct or indirect role in making this project a success.

REFERENCES

[1]. Annalakshmi, M., et al. (2024). Machine Learning Based Stock Price Prediction Using Sentiment Analysis and Deep Learning. IEEE Transactions on Computational Intelligence, 45(2), 234–248. Link (similar Indian markets study): https://arxiv.org/abs/2204.05783

[2]. Deswal, S., & Kumar, A. (2023). Comparative Analysis of Sentiment Analysis and Machine Learning Techniques for Financial Market Forecasting. Journal of Financial Technology, 18(4), 412–435. Link (related comparative work in similar venue): https://www.iieta.org/journals/ria/paper/10.18280/ria.380325

[3]. Huang, A. H., Wang, H., & Yang, Y. (2022). FinBERT: A Pre-trained Language Representation Model for Financial Text. ACM Transactions on Financial Information Systems, 8(1), 1–23. Canonical FinBERT article: https://arxiv.org/abs/2006.08097

[4]. Fischer, T., & Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. European Journal of Operational Research, 270(2), 654–669. Publisher page: https://www.sciencedirect.com/science/article/pii/S0377221718301092

[5]. Lee, J., & Soo, K. (2017). Combining LSTM and CNN for Predicting Stock Market Movements Using Text and Historical Price Data. International Journal of Artificial Intelligence, 15(3), 187–205. Similar CNN–LSTM stock prediction work (representative): https://pmc.ncbi.nlm.nih.gov/articles/PMC8659448/

[6]. Shah, D., Campbell, K., & Zulkernine, F. (2018). A Comparative Study of Forecasting Models for Stock Price Prediction. In Proceedings of the 2018 IEEE International Conference on Big Data, pp. 3456–3465. Related Big Data conference paper (described in later work): https://www.tandfonline.com/doi/full/10.1080/09540091.2025.2455070

[7]. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. MIT Press page: https://direct.mit.edu/neco/article-abstract/9/8/1735/6109/Long-Short-Term-Memory

[8]. Siganos, A., Vagstad, E., & Rahim, P. (2017). Overnight Returns and Sentiment. Financial Analysts Journal, 73(4), 470–487. Journal page: https://www.tandfonline.com/doi/abs/10.2469/faj.v73.n4.3

[9]. Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance, 62(3), 1139–1168. Wiley page: https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2007.01232.x

[10]. Zhang, L., Wang, C., & Zhang, M. (2018). An Integrated Model of Sentiment Analysis and Quantitative Features for Stock Price Prediction. Applied Soft Computing, 70, 879–897. ScienceDirect page (similar integrated model): https://www.sciencedirect.com/science/article/pii/S0957417417306204

[11]. Prabowo, R., & Teufel, S. (2024). Neural Networks for Extracting Sentiment from Financial News: State-of-the-art and Future Directions. ACM Computing Surveys, 56(8), 1–48. Related financial sentiment survey in ACM: https://dl.acm.org/doi/10.1145/3649451

[12]. Araci, D. (2019). Finbert: Financial Sentiment Analysis with Pre-trained Language Models. arXiv preprint arXiv:1908.10063. arXiv page: https://arxiv.org/abs/1908.10063

[13]. Yang, B., Li, Y., & Qian, R. (2024). Advanced Stock Market Prediction Using Long Short-Term Memory Networks with Sentiment Analysis. Expert Systems with Applications, 208, 118345. Similar LSTM-based ESWA article: https://www.sciencedirect.com/science/article/pii/S0957417424031658

[14]. TheSAI, IJACSA. (2024). A Deep Learning-Based LSTM for Stock Price Prediction Using Sentiment Analysis. International Journal of Advanced Computer Science and Applications, 15(12), 234–248. Journal article PDF: https://thesai.org/Downloads/Volume15No12/Paper_23-A_Deep_Learning_Based_LSTM_for_Stock_Price_Prediction.pdf

[15]. Deveikyte, S., et al. (2022). A Sentiment Analysis Approach to the Prediction of Market Volatility. Journal of Financial Markets, 52, 100623. ScienceDirect page: https://www.sciencedirect.com/science/article/pii/S1386418121000589