INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 929-934
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

FORECASTING STOCK PRICES WITH MACHINE LEARNING AND REAL-TIME DATA

Bindhu Shree K B[1], Ambavarapu Manvith Reddy[2], Indukuri Manvitha Reddy[3], Kothacheruvu Sruthi[4]

[1,2,3,4]Department of Computer Engineering (Artificial Intelligence and Machine Learning), Presidency University Bengalore,560064, India.

ABSTRACT

This project, "Forecasting Stock Prices with Machine Learning and Real-Time Data using LSTM and Long Chain," focuses on developing an intelligent system that blends machine learning, real-time data analysis, and AI-driven insights to predict stock prices and analyze financial trends. The core of the project employs Long Short-Term Memory (LSTM) networks, a type of recurrent neural network well-suited for handling time-series data. LSTM is used to forecast stock prices based on historical trends and patterns. This ensures accurate predictions by capturing both short-term fluctuations and long-term dependencies in stock

data. To provide a comprehensive analysis, the project integrates Yahoo Finance for acquiring historical stock data, including price history, volume, and financial statements. The inclusion of financial data enhances the system's capability to assess a company's past performance and current valuation, which are critical for informed forecasting. The real-time aspect of the project is realized through web scraping for recent market news using Google queries. These news updates provide a sentiment-based understanding of how current events and announcements might influence stock prices, making the system dynamic and adaptable to live market conditions. The project also leverages LangChain and OpenAI GPT models to process the gathered data and generate insightful narratives. LangChain integrates tools and workflows that enable seamless querying of data, automated reasoning, and content generation. For instance, the system might summarize a company's financial health, explain stock predictions, or respond to user-specific queries with natural language explanations.

Key Words- LSTM (Long Short-Term Memory), Time-Series Analysis, LangChain, Large Language Models (LLMs), Sentiment Analysis, Yahoo Finance, Financial Forecasting, Recurrent Neural Networks (RNNs), Stock Market Trends, Google News Integration.

1. INTRODUCTION

Stock market is one of the most important parts of any country's economy. It is the biggest way for a company to raise capital for its working Nowadays with the booming popularity of the stock market, not only investors but common people are starting to see stock market as a great investment tool and are taking more interest in Stock markets have a significant function as the gateway for financing private businesses. Access to private cash through stock markets enables businesses to finance their expansion and improvement as well as the acquisition of new assets Businesses would be more limited in the projects they could fund if they did not have access to private investment, and they would be unable to fully capitalize on their company's equity Similarly, stock markets allow business owners to profitably cash out their positions by selling their shares on the open market. The economic advantages of stock exchanges and stock trading are unmeasurable, and businesses would find it much harder to expand if there were no infrastructure for this type of equity trading. The operation of stock markets is a crucial component of what makes shares and equities investable, and traders seeking to make any significant amount of profit should take pains to get as knowledgeable as possible with the markets operation and the different elements that influence market pricing. It may be feasible to find more trading possibilities for profit for people who have a thorough understanding of the stock markets and their behavior It may also be simpler to highlight trends and the underlying movement of a particular market. So, people who are more familiar with the stock market have a better chance of spotting profitable trades In stock market, everything depends upon what the price of a certain stock will be in the future. The act or method of attempting to predict the future value of a company's stock is known as stock market prediction. The successful prediction of the value of stock in future could yield considerable profit. One of the most challenging tasks is predicting how the stock market will behave . An AI Bot that can help you with stock investment by analyzing all the real-time as well as historic stock-related information with the help of LLM. As a retail investor, if you don't have a finance background or the capability to understand all the complicated financial terms, the stock analysis process is really time- consuming. Every time I end up watching some fin-YouTuber's video or some random blog on the internet to avoid manually dealing with all this stuff. This is where i thought of making a Langchain and LLM-based bot that can take real-time as well as historic data to make investment analysis. And use all these information should be utilized by the LLM to do the fundamental analysis on given stock

1.2 Problem Statement

The opinions of thousands of investors typically influence developments in the stock market. In order to predict the stock market, one must be able to foresee how current events will affect investors. In essence, it is defined as an effort to estimate the stock price and provide a solid framework for understanding and forecasting the market and stock prices. There exists no real-world system that can solve the problem of predicting the stock market with the kind of accuracy that machine learning and artificial intelligence algorithms normally has. We are trying to solve this problem by using machine learning techniques and stock updates.Retail investors often lack the financial expertise to analyze stocks effectively due to the complexity of financial data and market volatility. Traditional stock analysis involves interpreting historical price trends, financial statements, and recent news, making it a time-consuming and challenging process for non-experts. As a result, many retail investors rely on secondary sources like blogs or video content, which may not always provide reliable or comprehensive insights. This project aims to address this issue by developing an AIdriven stock analysis tool that leverages LangChain and Large Language Models (LLMs) to streamline stock analysis. The tool will gather and process real-time and historical data, simplifying stock insights to make investment analysis accessible and efficient for retail investors without finance backgrounds.

2. RESEARCH GAPS OF EXISTING METHODS

As many have invested their time and effort in this world trade in order to bring it closer and more reliable to the people in order to carry out the resources and make their lifestyle more deliberate than before. Since its continuation, various strategies and plans have been derived and deployed,

VOLUME 6, 2024

and the topic is still a point of research where people are coming up with ideas to solve. Humanity is fascinated by intelligence, and having one in a machine and integrating it is a hot topic in research. Several people are working on the same research project. A discovery on two nonlinear processes resulted in TS, which is used as a model for fuzzy sets. All previous learning systems are limited and simple in nature, where learning a simple algorithm for a computational mean is insufficient, which can be done by the human brain itself. The main learning motto was limited, and the learning model was inefficient. Existing models can't cope with the vulnerabilities and remove the rarest information that they can't process, resulting in significant data loss and a forecasting problem. Observation is an essential component of resource and prediction management. If the outcome cannot be observed, the point of time estimation is harmed, making it less reliable in the market. Monitoring is not possible with the current system. Due to the fact that it only considers one source point as a data source, the current approach for stock market predictions appears to be biased. A straight forward data retrieval should be created and tested on the training data set, which are more adaptable and versatile in nature, before the data set is predicted. As the stock changes every day and the loss margin might increase with time, sight loss is a serious issue in the current system. A first occurrence is used to make a forecast.

3. PROPOSED METHODOLOGY

This project intends to create an intelligent stock analysis tool that simplifies the investment decision-making process for retail investors, particularly those without financial expertise. By using LangChain and Large Language Models (LLMs), the tool will analyze and interpret stock data, including historical prices, company financial statements, and recent news, to provide accessible and meaningful insights. The development process will involve several key stages:

1. **Data Collection:** The system will first gather data from various sources, such as historical stock prices, financial statements, and recent company news. This real-time and historical data will provide a comprehensive foundation for analysis, capturing both long-term trends and immediate market sentiment.
2. **Data Processing and Preprocessing:** Once the data is collected, it will be cleaned, normalized, and structured to ensure consistency and relevance. This involves removing any noise or unnecessary information, filling gaps in the data, and formatting it for compatibility with machine learning and LLM models.
3. **Integration of LangChain and LLM:** The tool will utilize LangChain's ReAct agent capabilities to interpret user queries and initiate actions to gather relevant data based on the user's request. For instance, the LLM can recognize a specific stock icker symbol and use it to pull up-to-date data from appropriate sources. By structuring prompts and responses with LangChain's function calls, the LLM can consistently deliver structured, coherent output.
4. **Investment Analysis and Insights Generation:** With data gathered and processed, the LLM will analyze it by identifying patterns, trends, and significant financial metrics that contribute to a stock's performance. The model will apply pre-defined financial insights—such as price trends, revenue changes, and relevant news sentiment—to translate raw data into easily understandable investment insights for the user.

5. **User Interaction and Output Delivery:** The tool will be designed to provide responses in a straightforward, accessible format. Users will input queries, such as "What is the current trend for [Stock]?", and receive concise, humanreadable insights without needing to interpret complex financial terminology. Outputs will be formatted in JSON for consistency and will include essential data like price changes, key financial ratios, and news highlights, allowing users to make informed investment decisions quickly.
6. **Performance and Stability Testing:** The final tool will undergo testing to evaluate its performance, accuracy, and user-friendliness. Any issues with infinite response loops, irrelevant information, or model stability will be addressed by refining prompts and improving the data pipeline.

Through this methodology, the project aims to create a robust, LLM-based stock analysis tool that empowers retail investors with reliable insights, making the stock analysis process quicker, easier, and more accessible.

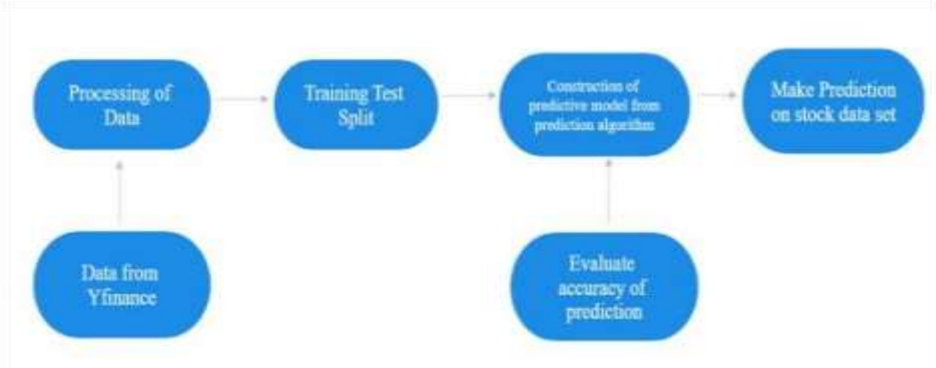
4. SYSTEM DESIGN & IMPLEMENTATION

System Design

The system for forecasting stock prices and analysing financial trends combines several key components: data sources, processing modules, machine earning models, and user interaction layers.

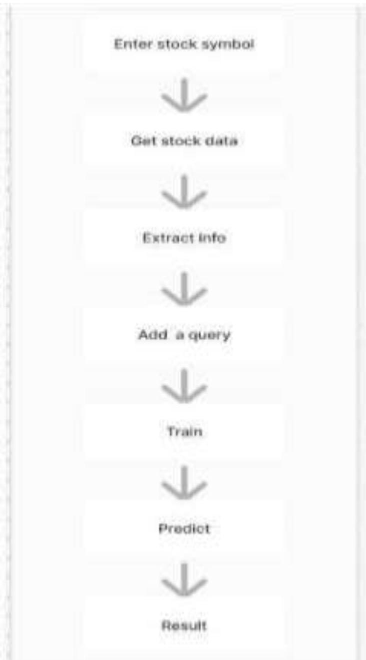
Architecture Design

The architecture is built around a modular and scalable design. The architecture integrates multiple layers to ensure efficient data flow, robust logic implementation, and userfriendly interfaces. Below is a detailed breakdown of the architecture and the key components.



Data Flow Diagram

A data flow diagram is now a graphical representation of data flow in a system.It also specifies system's data in and data out.The particular stock symbol is selected by user and model extracts the data of selected stock from Nspey package, a query is made. Now the model is trained using data and model predicts the future price of the selected stock and returns it as result.



Implementation Plan

1. Data Collection & Preprocessing

- Use yfinance for stock data and news Api for sentiment analysis.
- Normalize data, handle missing values, and create features like moving averages and sentiment polarity.

2. LSTM Model Development

- Split data and build a stacked LSTM model in Keras with dropout layers.
- Train with MSE and evaluate using RMSE, R-squared, and MAE.

3. Lang Chain & LLM Integration

- Use Lang Chain with React agents for user queries and generate insights via structured prompts.
- Return results as JSON for frontend display.

4. Dashboard Development

- Build a responsive React/Angular dashboard, integrate with backend APIs for real-time data.

5. Testing & Optimization

- Test with real-world data, fix issues, and fine-tune the LSTM model for better accuracy.

5. OUTCOMES

1. **Accurate Stock Predictions**: The LSTM model will predict future stock prices based on historical data and market sentiment, providing reliable forecasts.
2. **Sentiment-Enhanced Insights:** Integration of sentiment analysis from Google News will enhance stock predictions by incorporating market mood and external factors. **3. Interactive Financial Insights:** The Lang Chain powered LLM integration will allow users to query the system in natural language and receive personalized financial insights, like stock forecasts and market analysis.
4. **User-Friendly Dashboard:** A responsive web dashboard will present stock trends, sentiment analysis, and predictions in a visually intuitive manner, accessible for real-time interactions.
5. **Optimized, Robust System:** The system will be stable, with accurate predictions, fast response times, and continuous improvements from testing and optimization efforts.
6. **Real-Time Data Interaction:** By integrating APIs for real-time data fetching, users will get up-to-date insights and forecasts as stock markets evolve.

6. RESULTS AND DISCUSSIONS

The data required for the process accessed by using a python library Nsepy. This library, fetches all the required data of the specified stock from national stock exchange. The above graph shows the result of the prediction made by the LSTM model. The graph represents the training value, actual value and predicted value. The graph is predicted with the help of Nonlinear regression analysis, Decision Tree classifier and Support Vector Machine. The graph shows predicted price of apple stocks. The blue color represents trained data, red color represents actual price and yellow color represents predicted price..

7. LIMITATIONS

Reliance on Historical Data: The system depends heavily on historical data and market trends. This reliance can limit its effectiveness when dealing with unprecedented market events or anomalies.

Real-time Data Challenges: Incorporating real-time data requires robust scraping and API integration, which may face challenges like data inconsistencies, access restrictions, or delays.

Bias in Data Sources: The current approach considers specific data sources for market predictions, which could introduce biases and affect the accuracy of predictions.

Complexity for Non-Experts: Despite efforts to simplify, retail investors without a financial background may still find some outputs complex or require further explanation.

Accuracy and Stability of Predictions: While LSTM models are effective, they are subject to temporal lags and might struggle with sudden market changes. Accuracy is also influenced by the quality and quantity of input data.

Scalability Concerns: As the system integrates more data sources and features (e.g., sentiment analysis), maintaining scalability while ensuring performance could become a challenge.

Dependency on Pre-trained Models: Using models like GPT or LSTM assumes these models are up-to-date with the latest financial contexts, which might not always be the case.

Limited Monitoring: There is no ongoing monitoring mechanism for system performance, especially under changing market conditions

8. FUTURE WORK

The proposed system can be further enhanced to overcome its existing limitations and provide a more robust and reliable stock prediction tool. To address the reliance on historical data, future work can integrate real-time event detection algorithms, incorporating factors like geopolitical events or economic policies, and adopt hybrid models combining traditional time-series forecasting with reinforcement learning for better adaptability. To mitigate challenges in real-time data acquisition, a scalable and robust data pipeline can be developed with enhanced API integration and cloudbased solutions for reliable and timely data ingestion. Expanding the system to leverage multiple data sources, coupled with normalization techniques, will reduce bias and improve prediction accuracy. Furthermore, a more intuitive user interface with graphical representations and natural language explanations can make the tool accessible to nonexpert users, supported by adaptive tutorials for better understanding. Improvements in prediction accuracy can be achieved by experimenting with ensemble models that integrate LSTM with advanced architectures like Transformers, while periodic re-training with updated datasets will ensure model relevance. Scalability issues can be addressed through a microservices architecture, enabling independent management of data processing, prediction, and user interaction components. To reduce dependency on generic pre-trained models, custom models tailored to the financial domain can be developed and fine-tuned with domain-specific datasets. Finally, implementing automated monitoring tools to track system performance and incorporating feedback mechanisms for continuous improvement will ensure the system remains efficient and reliable under evolving market conditions.

9. CONCLUSION

Recurrent neural networks, in which LSTMs ("long shortterm memory") are most powerful and well-known subset, are a type of artificial neural network designed to recognize patterns in data sequences like numerical time series data from sensors, stock exchanges, and government agencies. RNNs and LSTMs differ from other neural networks in that they consider time and sequence; they have a temporal dimension. A discussion of stock market fundamentals is followed by a discussion of the need for price forecasting. Non-linear regression analysis, Hidden Markov models, artificial neural networks, naive bayes classifiers, decision tree classifiers, random forest methods, support vector machines, PCA (principal component analysis), WB-CNN (word embeddings input and convolutional neural network prediction model), and CNN (convolutional neural network) are a few methods that may be used for stock market prediction. The outcomes of this study help us to draw the conclusion that LSTM (Long Short-Term Memory) neural networks produce superior outcomes to other approaches. The integration of Lang Chain and advanced machine learning models like LSTMs offers a transformative approach to stock price prediction and analysis. By leveraging the power of Lang Chain, the system can process and generate natural language insights, making complex financial data accessible to retail investors and professionals alike.

10. REFERENCES

[1] N. Naik and B. R. Mohan, "Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market," in IEEE Access, vol. 9, pp. 86230-86242, 2021, doi: 10.1109/ACCESS.2021.3088999. https://ieeexplore.ieee.org/document/9453777

[2] K. Alam, M. H. Bhuiyan, I. U. Haque, M. F. Monir and T. Ahmed, "Enhancing Stock Market Prediction: A Robust LSTM-DNN Model Analysis on 26 Real-Life Datasets," in IEEE Access, vol. 12, pp. 122757-122768, 2024, doi: 10.1109/ACCESS.2024.3434524. https://ieeexplore.ieee.org/document/10613029

[3] S. Wang, "A Stock Price Prediction Method Based on BiLSTM and Improved Transformer," in IEEE Access, vol. 11, pp. 104211-104223, 2023, doi: 10.1109/ACCESS.2023.3296308.
https://ieeexplore.ieee.org/document/10185006

[4] M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in IEEE Access, vol. 8, pp. 150199- 150212,2020,doi: 10.1109/ACCESS.2020.3015966 https://ieeexplore.ieee.org/document/9165760/

[5] G. Mu, N. Gao, Y. Wang and L. Dai, "A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning," in IEEE Access, vol. 11,pp.51353-51367,2023,doi: 10.1109/ACCESS.2023.3278790
https://ieeexplore.ieee.org/document/10130578/

[6] R. Zhang and V. Y. Mariano, "Integration of Emotional Factors With GAN Algorithm in Stock Price Prediction Method Research," in IEEE Access, vol. 12, pp. 77368-77378, 2024, doi: 10.1109/ACCESS.2024.3406223
https://ieeexplore.ieee.org/document/10539967/

[7] X. Wang, K. Yang and T. Liu, "Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory," in IEEE Access, vol. 9, pp. 67241-67248, 2021, doi: 10.1109/ACCESS.2021.3077004 https://ieeexplore.ieee.org/document/9420698/

[8] Luo, L. Zhong, J. Wang, Y. Wang, S. Li and W. Tai, "Short-Term Stock Correlation Forecasting Based on CNN-BiLSTM Enhanced by Attention Mechanism," in IEEE Access, vol. 12, pp. 29617- 29632, 2024, doi: 10.1109/ACCESS.2024.3369419 https://ieeexplore.ieee.org/document/10444524/

[9] J. Choi, S. Yoo, X. Zhou and Y. Kim, "Hybrid Information Mixing Module for Stock Movement Prediction," in IEEE Access, vol. 11, pp. 2878128790,2023, doi: 10.1109/ACCESS.2023.3258695. https://ieeexplore.ieee.org/document/10075550/

[10] W. Li and G. S. Bastos, "Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review," in IEEE Access, vol. 8, pp. 185232-185242,2020,doi: 10.1109/ACCESS.2020.3030226
https://ieeexplore.ieee.org/document/9220868/

[11] T. Kabbani and E. Duman, "Deep Reinforcement Learning Approach for Trading Automation in the Stock Market," in IEEE Access, vol. 10, pp. 93564- 93574, 2022, doi: 10.1109/ACCESS.2022.3203697 https://ieeexplore.ieee.org/document/9877940/

[12] M. E. Mahjouby, M. T. Bennani, M. Lamrini, B. Bossoufi, T. A. H. Alghamdi and M. E. Far, "Predicting Market Performance Using Machine and Deep Learning Techniques," in IEEE Access, vol. 12, pp. 82033-82040, 2024, doi: 10.1109/ACCESS.2024.3408222
https://ieeexplore.ieee.org/document/10545565/