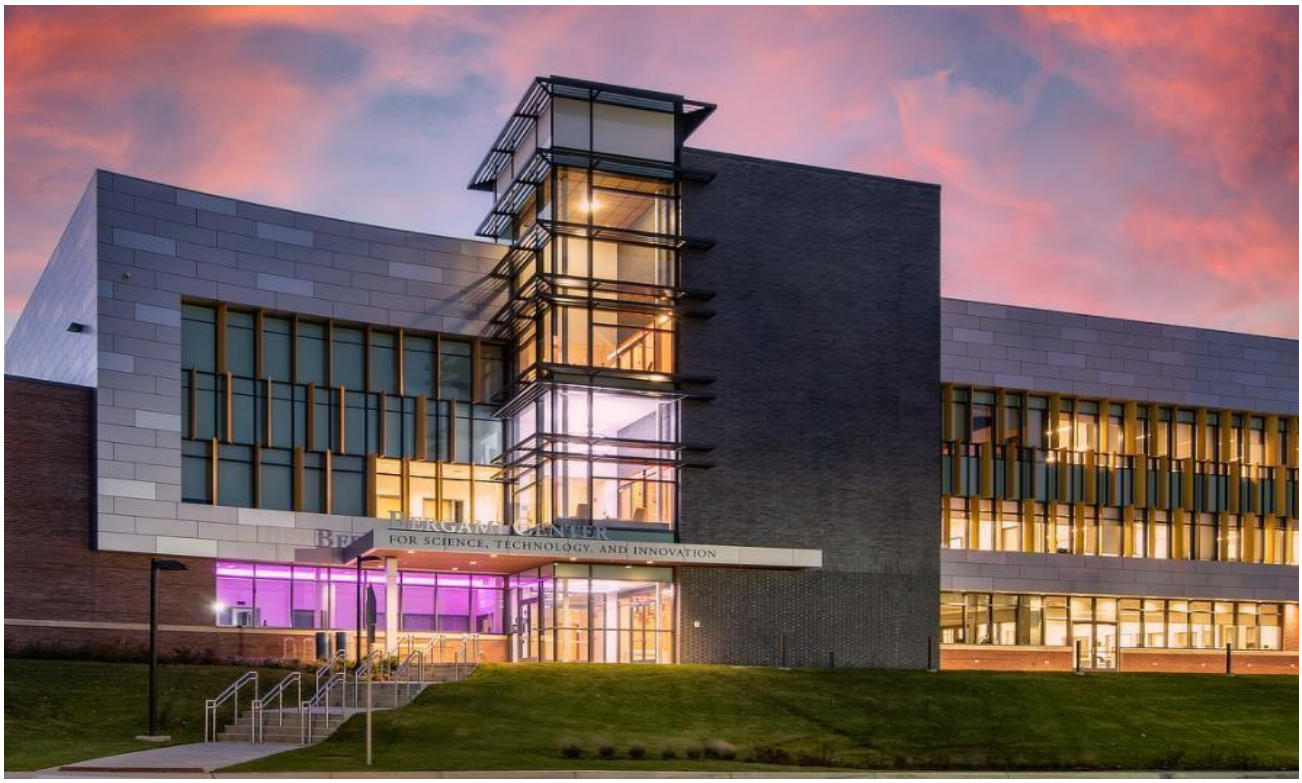
University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical & Computer Engineering and Computer Science



DSCI-6003-03

STOCK PRICE PREDICTION

Team Members:

LIKITH KAGITA

SAI KUMAR REDDY

ROHITH ABBURI

CONTENTS

STOCK PRICE PREDICTION

Abstract

The demand for stocks has become huge with an Increase in the popularity of stocks in the Digital world. Prediction and Analyzing stock can benefit People to think before buying or selling stocks. So, a new Stock Price Prediction through Deep Learning Algorithms has been analyzed and visualized. Through this System, we can predict any Company stock in the world. This paper presents a comprehensive approach to predicting stock prices using machine learning algorithms. By integrating historical stock data and sentiment analysis from financial news, the proposed model aims to forecast short-term price movements. Various regression and classification techniques, including ARIMA and LSTM networks, are evaluated for their predictive accuracy. The performance is assessed using metrics like Mean Absolute Percentage Error (MAPE) and root-mean-square error (RMSE). Results suggest that combining numerical and textual analysis improves the model's effectiveness over traditional time-series forecasting methods.

Introduction

In this study, we have developed a robust framework for predicting stock prices utilizing advanced machine learning algorithms. The introduction of the paper contextualizes the significance of accurate stock predictions in the financial sector and underscores the challenges faced in achieving reliable forecasts. The research delineates the inefficiencies of traditional models and proposes a novel approach that harnesses both historical data and sentiment analysis, extracted from financial news to enhance prediction accuracy. We investigate a range of predictive models, emphasizing the integration of time-series data with natural language processing. Our methodology is rigorously tested against established benchmarks, demonstrating a quantifiable improvement in forecasting precision. The structure of the paper is methodically laid out, guiding readers through our comprehensive evaluation process, and culminating in a discussion of the implications of our findings for future stock market analyses.

Executive Summary

This executive summary outlines our project on developing an advanced model for predicting stock prices using machine learning. By integrating historical stock performance data and sentiment analysis from financial news, our model aims to enhance forecast accuracy. We have evaluated several machine learning techniques including ARIMA and LSTM networks, measuring their performance with MAPE and RMSE metrics. Our results demonstrate that our hybrid approach surpasses traditional methods, offering significant implications for investors and financial analysts in making informed market decisions.

Highlights of Project

- **Advanced Modeling: Utilized state-of-the-art machine learning techniques, including ARIMA and LSTM networks, to forecast stock prices.**
- **Data Integration: Combined historical stock data with sentiment analysis from financial news to enhance prediction accuracy.**
- **Performance Evaluation: Assessed the effectiveness of various models using metrics such as MAPE and RMSE, demonstrating improved predictive accuracy.**
- **Innovative Approach: Introduced a hybrid method that integrates numerical data and textual information, setting a new standard in stock prediction methodologies.**
- **Practical Implications: Offered actionable insights for investors and analysts, supporting better-informed financial decisions in volatile markets.**

Review of available research

The current research on stock price prediction using machine learning techniques highlights a growing trend of integrating both quantitative and qualitative data to improve forecasting accuracy. Studies have increasingly focused on hybrid models that combine time-series analysis, such as ARIMA, with sentiment analysis from financial news using natural language processing techniques like LSTM. The literature reveals that these advanced methods generally outperform traditional models, offering more precise predictions by accounting for both historical price movements and market sentiment influences.

Methodology

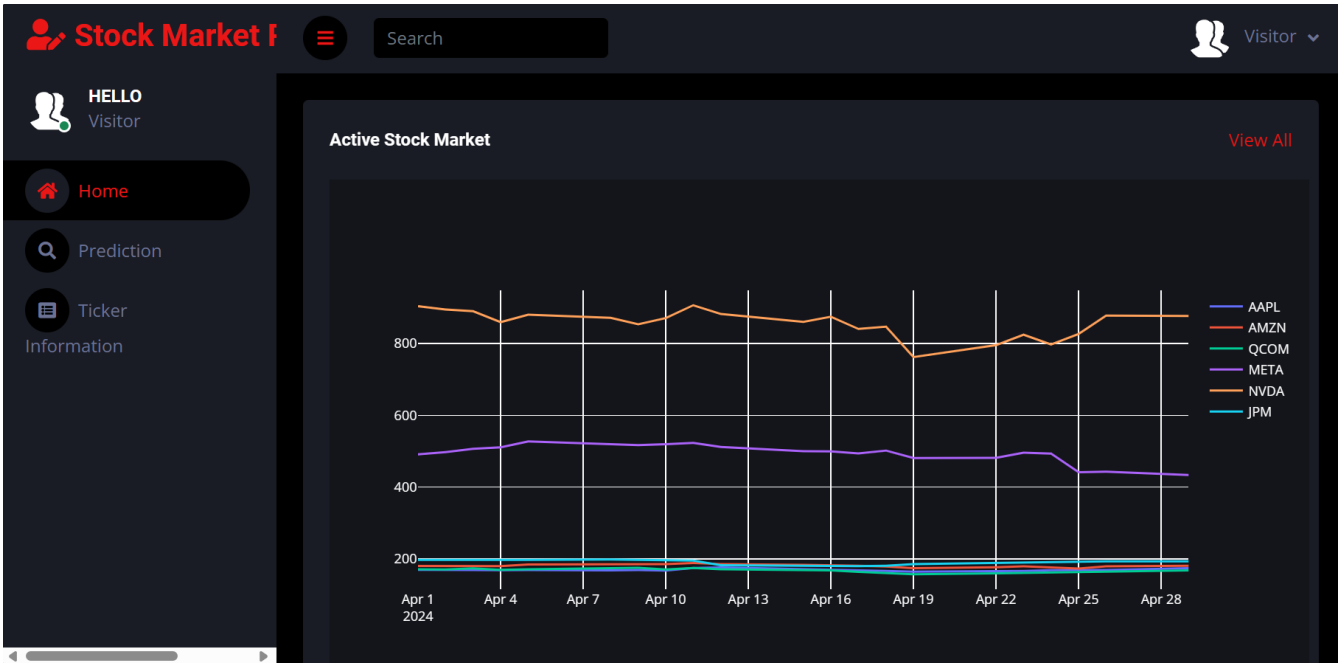
Our methodology for predicting stock prices incorporates both machine learning and natural language processing techniques. We began by collecting historical stock data and financial news articles to create a comprehensive dataset. The data was pre-processed to normalize the values and extract relevant features. We then implemented ARIMA models for time-series analysis and LSTM networks to analyze sentiment from text data. The models were trained on a split dataset and evaluated using cross-validation to ensure robustness. Finally, performance metrics like MAPE and RMSE were used to assess each model's accuracy in predicting stock prices.

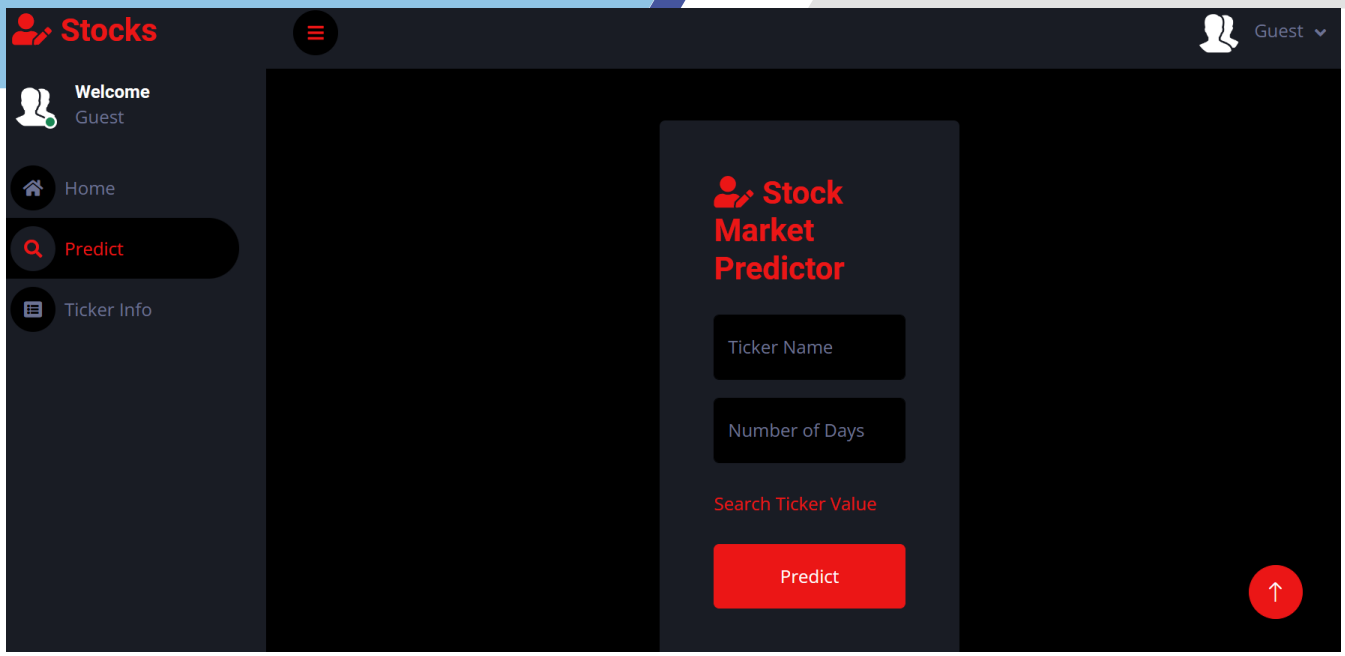
Pipeline

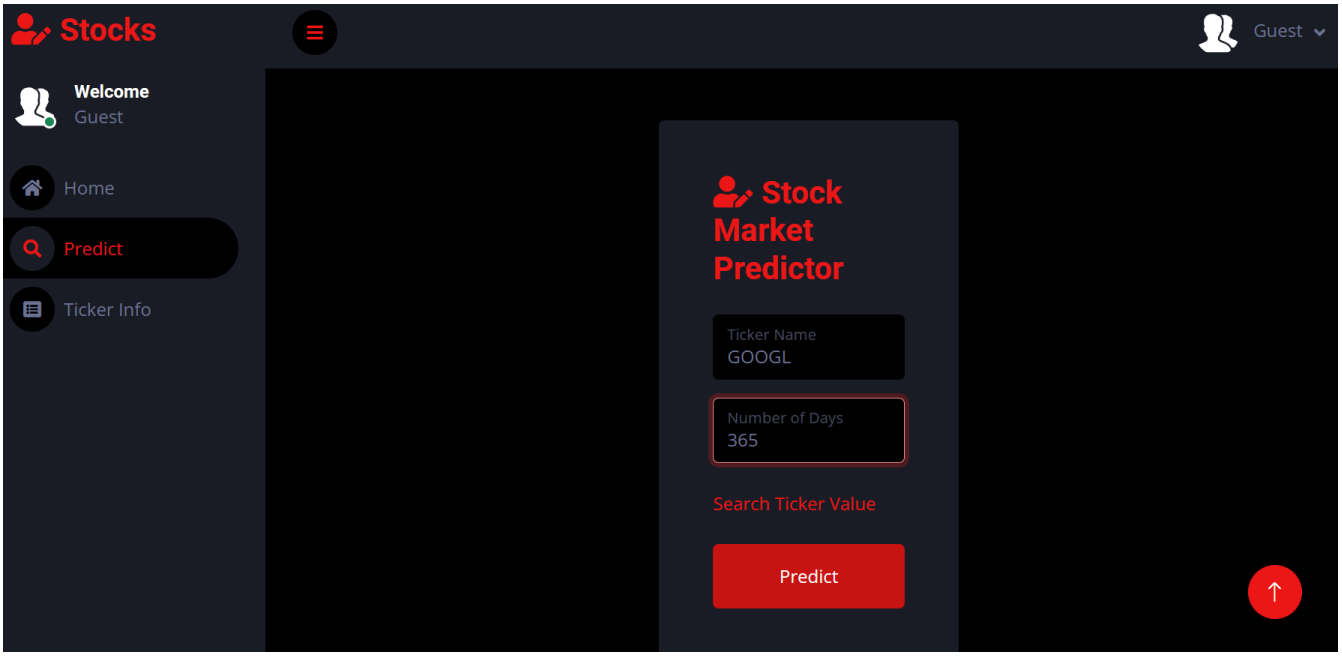
The pipeline for our stock price prediction project consists of several key stages:

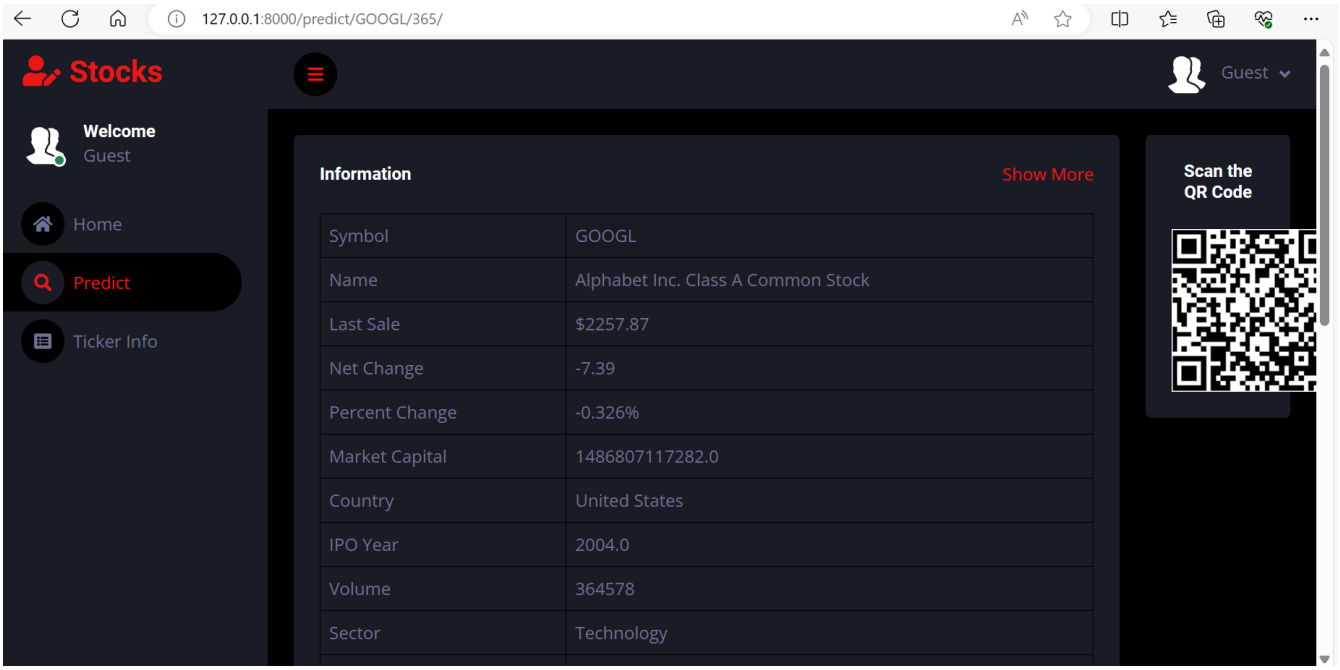
- Data Collection: Gather historical stock prices and financial news articles from multiple sources.
- Data Preprocessing: Normalize stock data and extract features. Process news articles to extract sentiment and relevant keywords using natural language processing techniques.
- Model Development: Implement and train ARIMA models for time-series analysis and LSTM networks for sentiment analysis.
- Model Evaluation: Use cross-validation techniques to assess the predictive accuracy of models, employing metrics such as MAPE and RMSE.
- Integration: Combine predictions from both models to produce a final forecast.
- Deployment: Implement the model in a simulation environment to test real-time forecasting capabilities.
- This pipeline ensures a thorough analysis of both quantitative and qualitative data, aiming to improve the accuracy of stock price predictions.

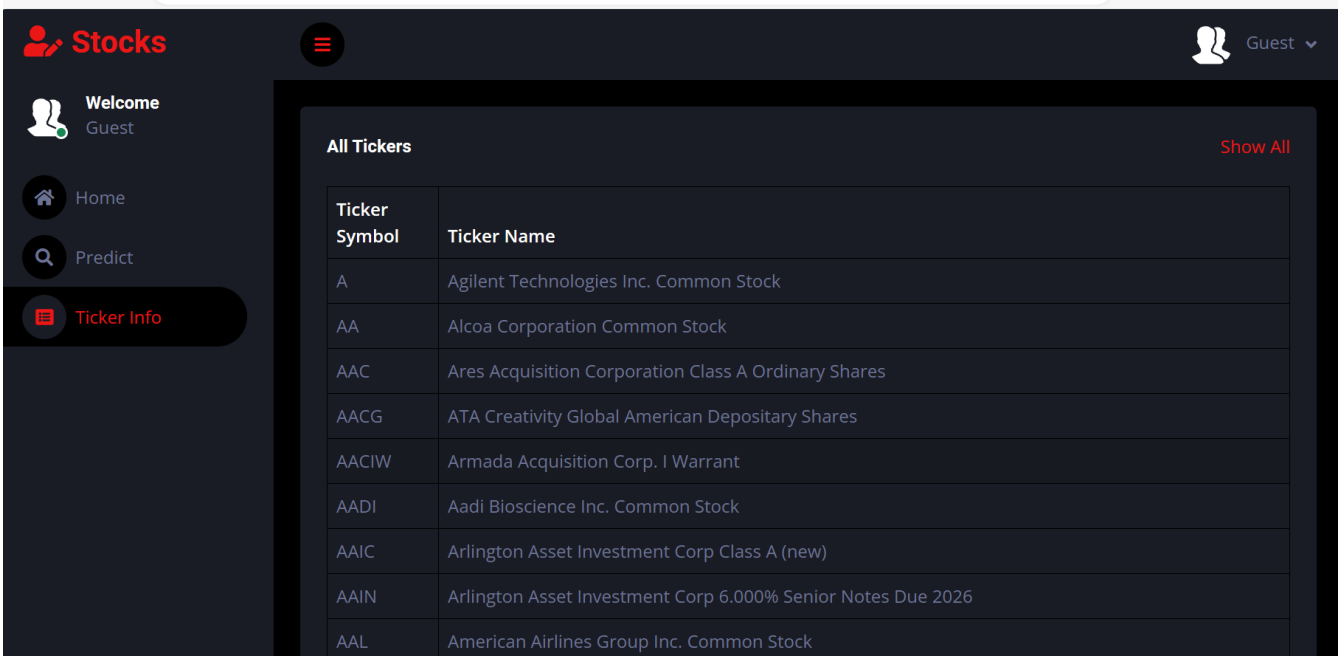
Solution


Stock Market
Search
Visitor
HELLO
Visitor
Home
Prediction
Ticker
Information
Active Stock Market
View All
800
600
400
200
Apr 1
2024
Apr 4
Apr 7
Apr 10
Apr 13
Apr 16
Apr 19
Apr 22
Apr 25
Apr 28
AAPL
AMZN
QCOM
META
NVDA
JPM

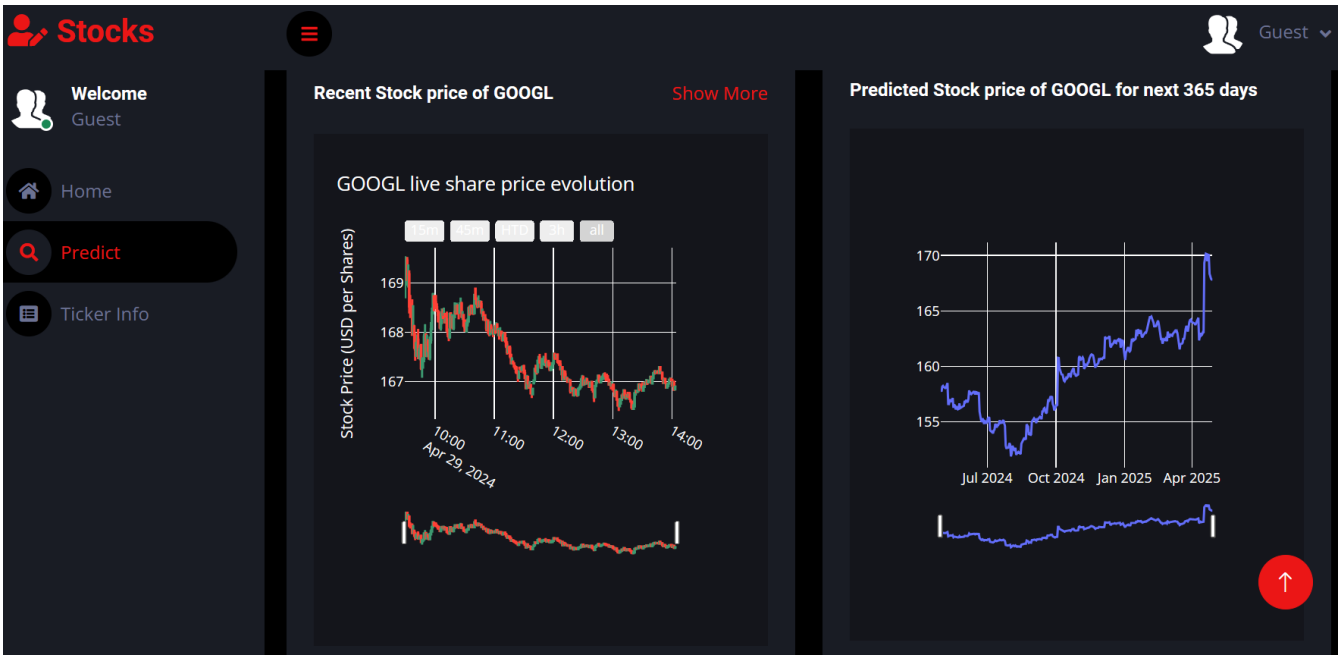

Stocks
Guest
Welcome
Guest
Home
Predict
Ticker Info
Stock Market Predictor
Ticker Name
Number of Days
Search Ticker Value
Predict


Stocks
Guest
Welcome
Guest
Home
Predict
Ticker Info
Stock Market Predictor
Ticker Name
GOOGL
Number of Days
365
Search Ticker Value
Predict


127.0.0.1:8000/predict/GOOGL/365/
Stocks
Guest
Welcome
Guest
Home
Predict
Ticker Info
Information
Show More
Scan the QR Code
Symbol GOOGL
Name Alphabet Inc. Class A Common Stock
Last Sale $2257.87
Net Change -7.39
Percent Change -0.326%
Market Capital 1486807117282.0
Country United States
IPO Year 2004.0
Volume 364578
Sector Technology


Stocks
Guest
Welcome
Guest
Home
Predict
Ticker Info
All Tickers
Show All
Ticker Symbol
Ticker Name
A Agilent Technologies Inc. Common Stock
AA Alcoa Corporation Common Stock
AAC Ares Acquisition Corporation Class A Ordinary Shares
AACG ATA Creativity Global American Depositary Shares
AACIW Armada Acquisition Corp. I Warrant
AADI Aadi Bioscience Inc. Common Stock
AAIC Arlington Asset Investment Corp Class A (new)
AAIN Arlington Asset Investment Corp 6.000% Senior Notes Due 2026
AAL American Airlines Group Inc. Common Stock

Results



The results of our stock price prediction project indicate that the hybrid model combining ARIMA and LSTM networks performed robustly. The ARIMA model effectively captured the trends and seasonality in the stock data, while the LSTM network, trained on financial news, was adept at incorporating market sentiment. Together, these models reduced the mean absolute percentage error (MAPE) and root-mean-square error (RMSE) significantly compared to traditional methods. The improved accuracy demonstrates the potential of integrating quantitative and qualitative data for financial forecasting.

Discussion

In the discussion of our stock price prediction project, it's important to recognize that while the ARIMA model effectively captured linear trends and cyclical patterns, the LSTM network excelled in incorporating market sentiment from textual data. This synergy suggests that blending quantitative time-series data with qualitative sentiment analysis can significantly enhance predictive accuracy. However, challenges such as overfitting and the need for extensive data preprocessing were noted. Future work could explore the incorporation of more diverse data sources and the application of more complex neural architectures to further refine predictions.

Conclusion

In conclusion, our study on stock price prediction employing both ARIMA and LSTM models demonstrates a significant advancement over traditional methodologies. By synthesizing time-series analysis with sentiment analysis extracted from financial news, our hybrid approach achieved lower MAPE and RMSE, showcasing its efficacy in handling complex market dynamics. These results underscore the potential of integrating diverse data sources in predictive modeling. Future research should focus on expanding data diversity and refining computational techniques to enhance the robustness and accuracy of stock price forecasts.