



IT5602 - DATA SCIENCE AND ANALYTICS PROJECT REPORT

STOCK MARKET SENTIMENT ANALYSIS

BY:

JANESH S

2022506002

KARTHICKKANNAN A

2022506023

Introduction

Stock markets are highly volatile and influenced by numerous factors including news sentiment, financial indicators, and market trends. With advancements in deep learning and time series forecasting, there is an opportunity to build a system that not only predicts stock prices but also incorporates real-time sentiment analysis from financial news headlines.

This project combines statistical (ARIMA), machine learning (Linear Regression), and deep learning (LSTM) models along with sentiment analysis using NLP to provide a hybrid stock market analysis and recommendation platform.

Problem Statement

The project tackles the dual challenge of forecasting stock prices and integrating news-based sentiment for enhanced decision making. It aims to address the need for a predictive system that is robust against market noise and can suggest trading actions based on both price trends and public sentiment.

Objectives

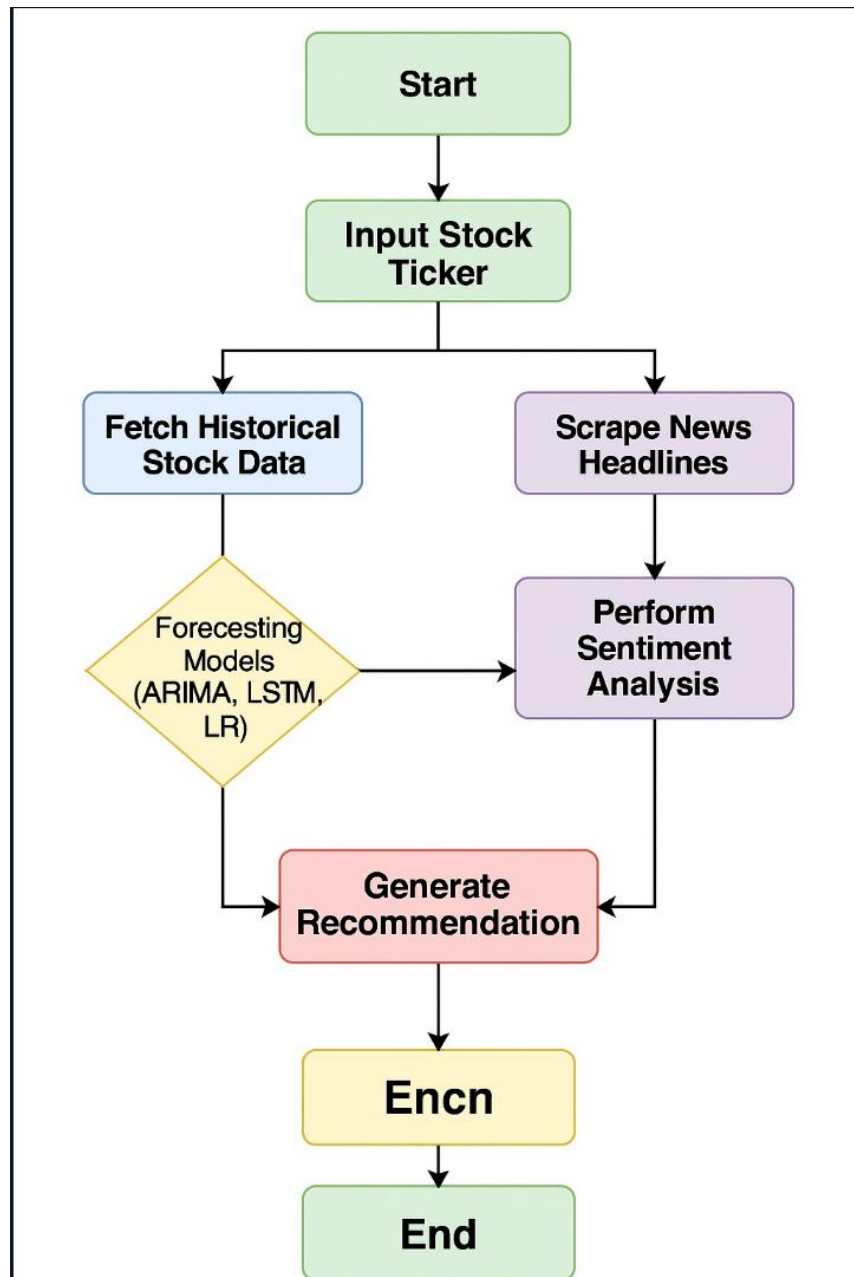
The main objectives of this project are as follows:

- ❖ To collect and preprocess historical stock data.
- ❖ To extract news headlines and compute sentiment polarity.
- ❖ To apply **ARIMA, LSTM, and Linear Regression** models for time-series forecasting.
- ❖ To compare models based on **RMSE** and visualization of predictions.
- ❖ To derive trading recommendations based on forecast and sentiment score.
- ❖ To visualize sentiment distribution and model predictions.

Tools and Technologies Used

- ❖ **Python 3.x** – Primary language for implementation.
- ❖ **Pandas** – Data handling and preprocessing.
- ❖ **NumPy** – Numerical operations and matrix transformations.
- ❖ **Matplotlib & Seaborn** – For visualizations and plotting.
- ❖ **Scikit-learn** – For Linear Regression and scaling.
- ❖ **Keras (TensorFlow backend)** – For building LSTM model.
- ❖ **Statsmodels** – For ARIMA time-series modeling.
- ❖ **TextBlob** – For sentiment analysis using polarity scoring.
- ❖ **BeautifulSoup** – Web scraping news from FinViz.
- ❖ **yFinance & Alpha Vantage** – APIs to retrieve stock data.
- ❖ **Streamlit** – For building the user interface.
- ❖ **ngrok** – To expose the local app for remote access

Usecase Diagram:



Methodology

1. Data Collection & Preprocessing

- Retrieve stock data using yFinance or Alpha Vantage.
- Clean and format columns: Date, Open, High, Low, Close, Volume.
- Filter out invalid entries and handle missing values.

2. Sentiment Analysis

- Scrape recent headlines from FinViz for a given stock ticker.
- Apply TextBlob to compute sentiment polarity scores.
- Categorize sentiment as Positive, Negative, or Neutral.

3. Modeling

- **ARIMA Model:**
 - Forecast short-term price movements using autoregressive and moving average terms.
 - Evaluate using RMSE and plot actual vs predicted values.
- **LSTM Model:**
 - Deep learning model with 4 hidden layers, trained on closing prices.
 - Normalize input, reshape for sequence learning, and forecast future prices.
- **Linear Regression Model:**
 - Predicts closing price based on previous trend using regression.
 - Compared against ARIMA and LSTM based on RMSE.

4. Recommendation Engine

- Combine latest predicted price and mean of 7-day forecast.
- Use sentiment polarity to adjust recommendation:
 - If price is undervalued and sentiment is positive → Recommend BUY
 - Else → Recommend SELL

System Architecture

The system comprises the following components:

- ❖ **Data Handler:** Fetches and preprocesses stock data.
- ❖ **Sentiment Analyzer:** Scrapes and processes headlines into sentiment scores.
- ❖ **ARIMA Engine:** Performs statistical price prediction.
- ❖ **LSTM Engine:** Trains and forecasts with deep learning.
- ❖ **Linear Regressor:** Provides comparison baseline.
- ❖ **Decision Engine:** Combines sentiment and forecast for recommendations.
- ❖ **UI Layer:** Interactive front-end built using Streamlit.

Results and Observations

- ❖ LSTM provided smoother and more accurate trends compared to ARIMA.
- ❖ ARIMA handled short-term fluctuations well but struggled with nonlinear
 - patterns.
- ❖ News sentiment aligned well with market movement in test cases.
- ❖ Combining prediction with sentiment yielded more robust trading advice.

Evaluation Metrics

- ❖ **RMSE (Root Mean Square Error)** – For each model to evaluate prediction accuracy.
- ❖ **Sentiment Polarity Mean** – Used to classify news sentiment.
- ❖ **Visual Accuracy** – Line charts helped visualize forecast accuracy.

Visualizations

- ❖ ARIMA, LSTM, Linear Regression model predictions (Actual vs Predicted).
- ❖ Sentiment Pie Chart: Distribution of Positive, Negative, Neutral headlines.
- ❖ Forecast plot of 7-day predicted stock prices.

Input and Output

Stock Price Prediction with ARIMA, LSTM & Linear Regression

Select stock ticker:

TSLA

Choose action:

☐ Load Saved Data

☒ Fetch New Data

Load / Fetch Data

Fetching data for TSLA from 2023-05-18 to 2025-05-18...

Data saved to D:\Stock-Prediction & Sentiment analysis project\UITSLA.csv

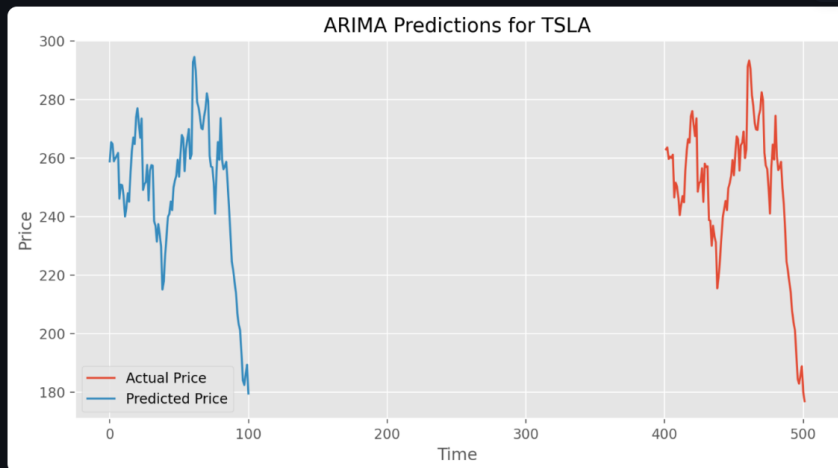
	Date	Close	High	Low	Open	Volume
		TSLA	TSLA	TSLA	TSLA	TSLA
496	2023-05-24 00:00:00	182.9	184.22	178.22	182.23	137605100
497	2023-05-23 00:00:00	185.77	192.96	185.26	186.2	156952100
498	2023-05-22 00:00:00	188.87	189.32	180.11	180.7	132001400
499	2023-05-19 00:00:00	180.14	181.95	176.31	177.17	136024200
500	2023-05-18 00:00:00	176.89	177.06	172.45	174.22	109520300

ARIMA MODEL

☐ Run ARIMA model

Running ARIMA model for TSLA...

Running ARIMA model for TSLA...



ARIMA RMSE: 7.7824

ARIMA next day predicted price: 201.12

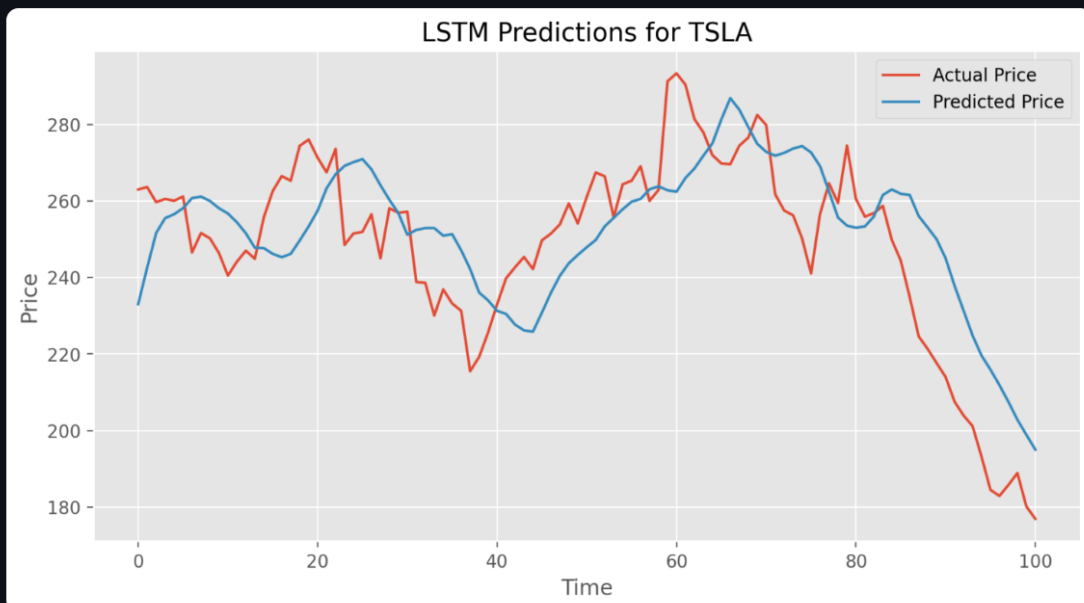
ARIMA 7-day forecast: [201.12417861 193.00211415 184.03737404 182.45004647 186.09965352 189.38212214 179.52877584]

LSTM MODEL

☐ Run LSTM model

Running LSTM model for TSLA...

Running LSTM model for TSLA...



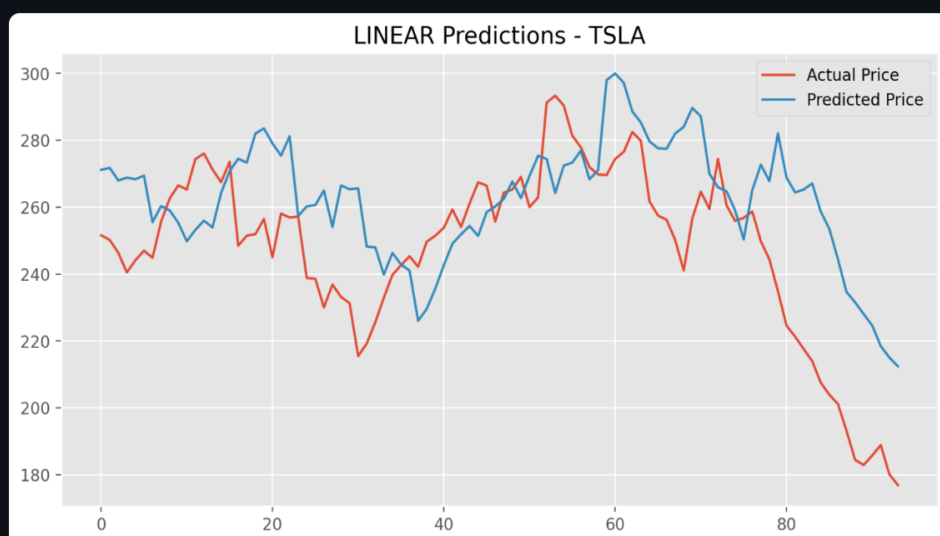
LSTM RMSE: 17.1157

LSTM next day predicted price: 233.03

LSTM 7-day forecast: [233.03203 242.11327 249.88963 251.65445 249.84158 248.07079 247.00441]

LINEAR REGRESSION MODEL

☐ Run Linear Regression model



Linear Regression next day predicted price: 204.84

Linear Regression 7-day forecast: [204.84411668 196.58088233 195.08969348 197.81562183 200.75998503 192.46825791 189.38141641]

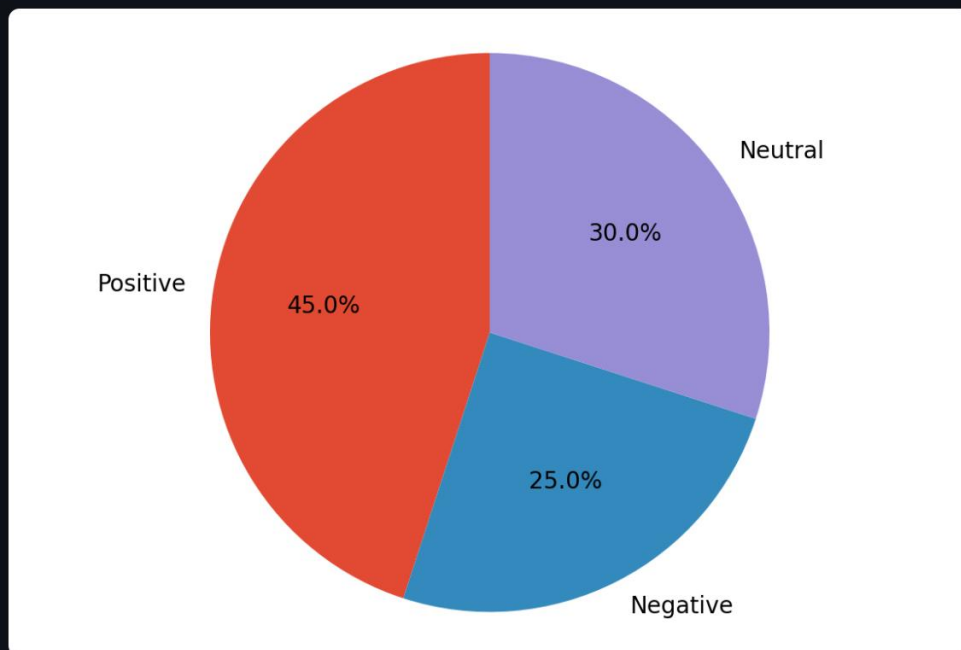
Mean forecasted price: 196.71

Linear Regression RMSE: 25.1954

SENTIMENT ANALYSIS :

☐ Run Sentiment Analysis (FinViz.com) and Recommendation

Positive Tweets : 9, Negative Tweets : 5, Neutral Tweets : 6



News Polarity: Overall Positive

Conclusion

This project showcases a powerful approach to stock market prediction by combining deep learning, time-series modeling, and real-time sentiment analysis.

The hybrid model helps in making informed investment decisions backed by data-driven forecasts and market sentiment interpretation.

