World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Effective market hypothesis using machine learning

Niharika Pilligundla, Yagnesh Chimpiri *, Uday Kiran Kandukoori, Vaishnav Teja Jonnalagadda and Shiva Shashank Pagadala

Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India.

World Journal of Advanced Research and Reviews, 2025, 25(02), 553-560

Publication history: Received on 25 December 2024; revised on 01 February 2025; accepted on 04 February 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.2.0363

Abstract

The Effective Market Hypothesis is designed to predict stock prices using Long Short Term Memory networks (LSTM), a kind of recurrent neural networks that are particularly suitable for modeling sequential and time series data. The LSTM model is designed to capture complex patterns, trends and potential market behaviours and, using historical stock price data and technical indicators, to model them, something that is often difficult in traditional statistical methods. Challenges such as market volatility, nonlinear dynamics and external factors that can have a significant impact on stock prices are also addressed. Model training and evaluation are then carried out rigorously, and the potential of LSTMs to increase the accuracy and reliability of stock price predictions is explored to deeper insights into market movements. In volatile market conditions, these findings are meant to help improve investment strategies, risk management practices, and the understanding of how machine learning techniques can be applied to financial forecasting.

Keywords: Effective Machine Learning; Long Short Term Memory; Recurrent Neural networks; Model training

1. Introduction

In today's rapidly evolving financial landscape, accuracy of predicting stock prices is prominent for investors, traders, and financial institutions seeking to make accurate decisions and manage risks efficiently. However, existing prediction models often face difficulty to address the complex and dynamic nature of the stock market, mainly caused due to major factor including economic trends, geopolitical events, and investor sentiment can lead to significant price fluctuations. This limitation effect the ability to accurately predict future stock movements and optimize investment strategies.

Our project, Effective Market Hypothesis using Machine Learning, solve these challenges by employing advanced machine learning techniques, specifically Long Short-Term Memory (LSTM) networks, to predict stock prices with higher accuracy. LSTM, a type of recurrent neural network (RNN), is well suited for handling time series data, enabling the model to capture complex patterns and trends in historical stock prices. By incorporating technical indicators and other relevant market factors, the LSTM model can predict future price movements, even in the face of market volatility.

The application of this project is extensive. In finance, it benefits investors and traders in making data-driven decisions with less risk and more return. In portfolio management, LSTM based predictions can assist in optimizing asset allocation by identifying market trends. Furthermore, this project is relevant to the larger field of machine learning in finance by giving a picture of how well deep learning algorithms work for predicting stock prices. This research therefore aims to improve investment strategies, risk management, and market behaviour understanding, by enabling more accurate predictions. Thereby contributing to the financial industry decision making process.

* Corresponding author: C Yagnesh

2. Related Work

Forecasting stock value has been a tough cookie for finance and machine learning over recent years a different types of methods have been tried to surpass these with traditional methods like time series analysis and other estimation models being prominent for a while but machine learning evolved especially along with deep learning techniques models like recurrent neural networks rnns and long short-term memory lstm networks have really taken center stage due to their ability to predict the complex trends that show up in data over a period of time in earlier attempts at stock value prediction machine learning models like support vector machines svms were utilized as seen in the work of zhang et al 2018 who focused over past price data and technical estimators while the results were somewhat promising these models struggled to effectively capture the long-term patterns and more intricate trends that are crucial for estimating stock prices other machine learning techniques like decision trees and other techniques chen et al 2019 were also tested but they often could not quite match up to models designed for time-dependent data like lstms.

The real breakthrough came when lstm networks started being applied to stock prediction lstms a special type of rnns are great at identifying patterns in data that occur over long periods making them a perfect fit for time-series tasks like predicting stock prices for example fischer and krauss 2018 used lstms for estimating stock price movements for different companies and their model outperformed older techniques like lstm and even other machine learning models similarly jiang et al 2020 applied lstms using just historical stock prices and saw significant improvements compared to traditional approaches however even with lstms showing great promise there are still some challenges one of the key issues that come up in recent studies is the black-box nature of lstm models.

While they are fantastic at making accurate estimations understanding why the model makes a particular prediction is not always straightforward which can be a drawback when trying to trust and interpret the results liu et al 2021 moreover most studies up until now have focused primarily on using just previous market price data but in the real world stock prices can be influenced by so much more things like market sentiment financial news and social media chatter karim et al 2020 a few studies have tried incorporating additional data sources like xu et al 2022 who added sentiment analysis from financial news to improve stock predictions but there is still a lot of untapped potential in using multi-source data in our study we aim to fill those gaps by enhancing lstm based models with not just historical stock prices but also technical estimators and sentiment analysis from sources like social media and news articles by incorporating these additional data points we hope to improve overall accuracy and at the same time make the models decisions easier to understand and interpret.

3. Existing System

Many existing systems rely on traditional statistical models like arima autoregressive integrated moving average or machine learning algorithms like random forest and support vector machines while these methods work well in certain situations they tend to face difficulty when it comes to capturing the time-based relationships and complex non-linear patterns in financial data on top of that these systems often require a lot of manual feature engineering and specialized knowledge.

Making them less adaptable and harder to scale as financial markets become more complex and fast-changing the need for manual data processing and pre-set assumptions about how data behave makes it difficult for these models to keep up with shifting market conditions furthermore these systems often miss the subtle connections between different factors that influence stock prices which means their predictions are not as accurate especially during times of high volatility or unexpected market changes.

4. Proposed Model

The proposed system leverages an enhanced long short-term memory lstm model tailored for stock price prediction lstm is well-suited for capturing sequential dependencies and complex patterns in time-series data making it ideal for financial forecasting the system integrates advanced preprocessing techniques including log returns transformation and outlier removal to ensure data quality feature engineering enriches the input with key indicators like moving averages rsi and external factors such as news sentiment and macroeconomic variables providing a holistic view of market trends to boost model performance.

The architecture includes stacked lstm layers for capturing deeper patterns attention mechanisms for highlighting important inputs and regularization techniques like dropout and l2 regularization to mitigate overfitting the system is trained using optimized loss functions and adaptive learning rate schedules to improve convergence evaluation combines traditional metrics rmse mape with back testing to assess real-world trading applicability ensuring the model is both accurate and practical for deployment.

5. Methodology

The methodology involves collecting historical stock data, preprocessing it by handling missing values and normalizing features, and performing feature engineering with indicators like moving averages and RSI. A stacked LSTM model is then trained on this data using backpropagation and optimized through hyperparameter tuning. The model is evaluated using metrics like MSE and RMSE, followed by deployment in a real-time system with continuous updates for accuracy.

Description of the Machine Learning Algorithms and Techniques Chosen: Libraries imported are:

Import all the required libraries

```
import pandas as pd
import datetime as dt
from datetime import date
import matplotlib.pyplot as plt
import yfinance as yf
import numpy as np
import tensorflow as tf
```

Figure 1 Libraries imported

5.1. Data Collection

Gather historical stock price data from sources like Yahoo Finance, Alpha Vantage, or Quandl. The data includes stock prices, trading volumes, and financial indicators. External factors like news sentiment and economic trends may also be considered.

5.1.1. Data Preprocessing

Handle missing values, remove outliers, and normalize features to ensure consistency. This step helps prevent biases and ensures that all features contribute equally to model learning. Data is then structured into a format suitable for time-series analysis.

5.2. System Architecture

The system architecture consists of data collection, preprocessing, feature engineering, model training, evaluation, and deployment. A Long Short-Term Memory (LSTM) model is used to process time-series stock data, capturing temporal dependencies for accurate predictions. The trained model is integrated into a real-time system, continuously updated with new data for improved forecasting.

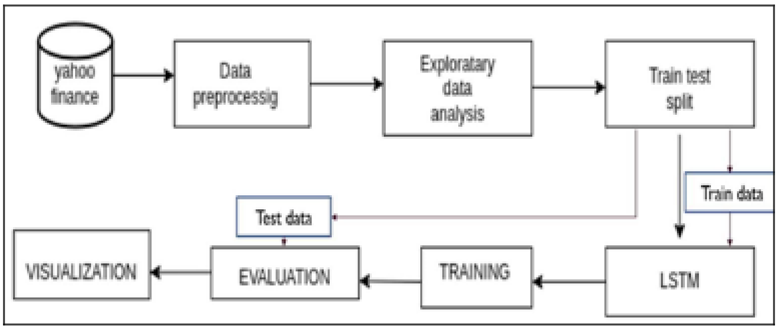

Figure 2 System Architecture

It illustrates how input data flows through various processing layers, resulting in meaningful outputs.

5.2.1. System Architecture Components

Input Module

Historical stock data (prices, trading volumes, financial indicators)

External factors (news sentiment, economic indicators).

Feature Engineering

Extract technical indicators (moving averages, RSI, MACD)Incorporate sentiment analysis and external market trends

Evaluation & Optimization

Validate model performance using metrics MSE and RMSE. Tune hyperparameters to improve accuracy

Prediction

Assigns confidence levels to the detected emotions, providing a measure of prediction reliability.

Output

Predicted stock prices for the next time period Visualization of predicted vs. actual stock trends

Deployment & Monitoring

Deploy model into a real-time or near-real-time system.

Continuously update and retrain using new data for better accuracy.

5.3. Model Development

Once the model achieves satisfactory performance, it can be deployed in a production environment to provide real-time predictions or periodic forecasts.

5.3.1. Define LSTM cell Architecture:

The Efficient Market Hypothesis (EMH) suggests that asset prices fully reflect all available information at any given time. According to this theory, it's impossible to "beat the market" consistently through expert stock picking or market timing because prices always incorporate and reflect all relevant data.

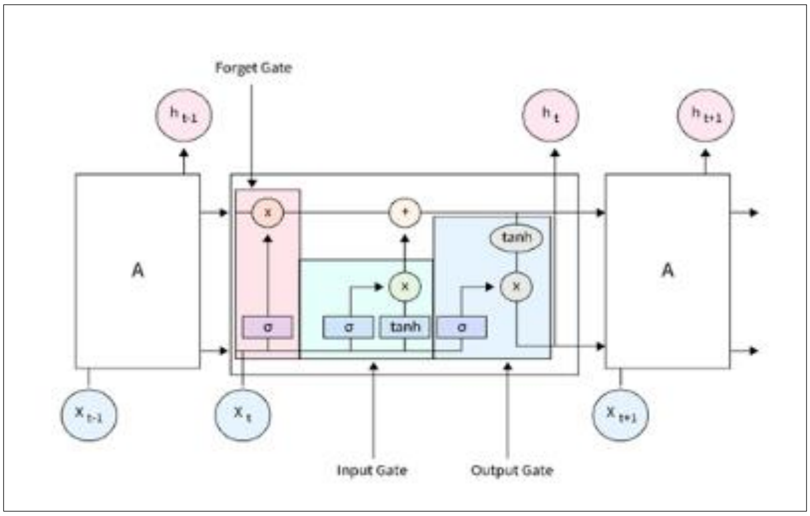

Figure 3 LSTM Cell Architecture

5.3.2. Validation

The validation of the Efficient Market Hypothesis (EMH) is often tested through empirical studies that analyze stock price movements, trading strategies, and market anomalies. Many studies support EMH by showing that stock prices follow a random walk, meaning past prices do not predict future movements, aligning with the idea that markets incorporate all available information.

```
from tensorflow.keras.layers import Dense, Dropout, LSTM
from tensorflow.keras.models import Sequential

model = Sequential()
model.add(LSTM(units = 50, activation = 'relu', return_sequences=True
              ,input_shape = (x_train.shape[1], 1)))
model.add(Dropout(0.2))


model.add(LSTM(units = 60, activation = 'relu', return_sequences=True))
model.add(Dropout(0.3))


model.add(LSTM(units = 80, activation = 'relu', return_sequences=True))
model.add(Dropout(0.4))


model.add(LSTM(units = 120, activation = 'relu'))
model.add(Dropout(0.5))

model.add(Dense(units = 1))
```

Figure 4 Model Building

5.3.3. Training

Training a model to validate the Efficient Market Hypothesis (EMH) typically involves using machine learning techniques, such as LSTM networks, to predict stock prices or returns. The process includes collecting historical financial data, preprocessing it (e.g., normalization, feature selection), and training the model on past price movements.

Training the model

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error', metrics = ['MAE'])
model.fit(x_train, y_train, validation_data = (x_test, y_test) ,epochs = 50)

Epoch 1/50
43/43 [==============================] - 41s 806ms/step - loss: 0.0021 - MAE: 0.0296 - val_loss: 4476.0820 - val_MAE: 64.940
Epoch 2/50
43/43 [==============================] - 32s 748ms/step - loss: 0.0018 - MAE: 0.0282 - val_loss: 4482.5337 - val_MAE: 64.990
Epoch 3/50
43/43 [==============================] - 34s 785ms/step - loss: 0.0014 - MAE: 0.0252 - val_loss: 4479.4561 - val_MAE: 64.968
Epoch 4/50
43/43 [==============================] - 34s 783ms/step - loss: 0.0015 - MAE: 0.0258 - val_loss: 4478.7212 - val_MAE: 64.963
Epoch 5/50
43/43 [==============================] - 35s 814ms/step - loss: 0.0016 - MAE: 0.0269 - val_loss: 4477.4165 - val_MAE: 64.953
Epoch 6/50
43/43 [==============================] - 32s 742ms/step - loss: 0.0017 - MAE: 0.0274 - val_loss: 4482.0625 - val_MAE: 64.990
Epoch 7/50
43/43 [==============================] - 34s 790ms/step - loss: 0.0022 - MAE: 0.0316 - val_loss: 4483.4443 - val_MAE: 64.997
```

Figure 5 Model Training

5.3.4. Testing

Testing the Efficient Market Hypothesis (EMH) using a trained model, such as an LSTM, involves evaluating its predictive power on unseen stock price data. After training, the model is tested on a separate dataset to check whether it can predict future price movements with consistent accuracy.

5.4. Real-Time Data Processing

This step ensures the system operates effectively in a real-time environment.

5.4.1. Cleaning

Handle missing values (using techniques like interpolation or deletion), remove outliers, and ensure data consistency.

Normalization/Scaling: Normalize or standardize the stock price data to a consistent scale, such as using Min-Max scaling or Z-score normalization. This helps the LSTM model to learn better from the data

5.4.2. Feature Extraction

Identify relevant features that contribute significantly to predicting stock prices. Technical indicators (moving averages, RSI), fundamental factors (earnings, dividends), and market sentiment can be crucial.

5.5. Stock price prediction

This step involves predicting stock prices based on the extracted features.

5.5.1. Model Prediction

Apply the trained model to new data to make predictions on future stock prices. Continuous monitoring and retraining may be necessary to adapt the model to changing market conditions.

5.6. Visualization

Delivering results in a user-friendly manner is crucial for system usability.

5.6.1. Real-Time Display

Developed a graphical user interface (GUI) using tool like Stream lit. Displaying stock trends of selected individual companies with opted number of days prediction.

6. Results and Discussion

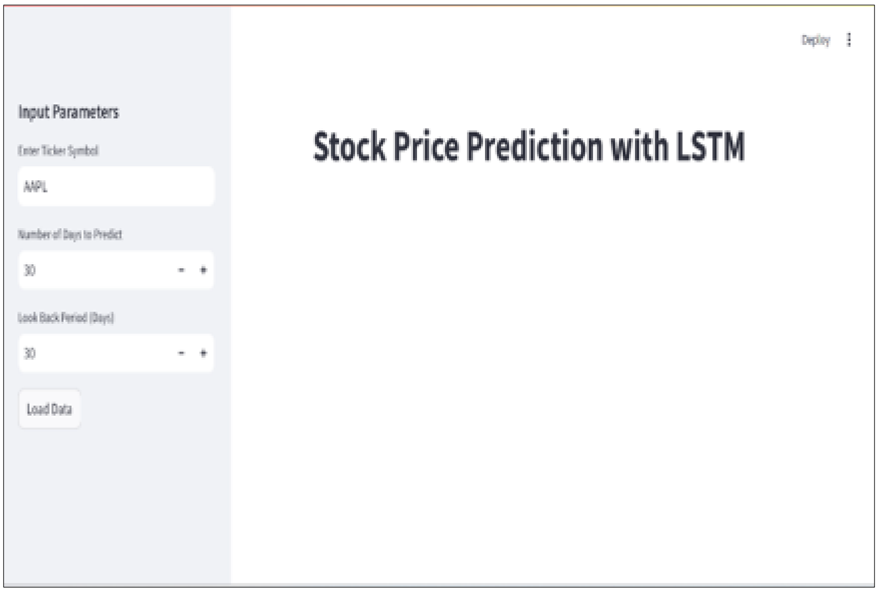

Figure 7 User Interface

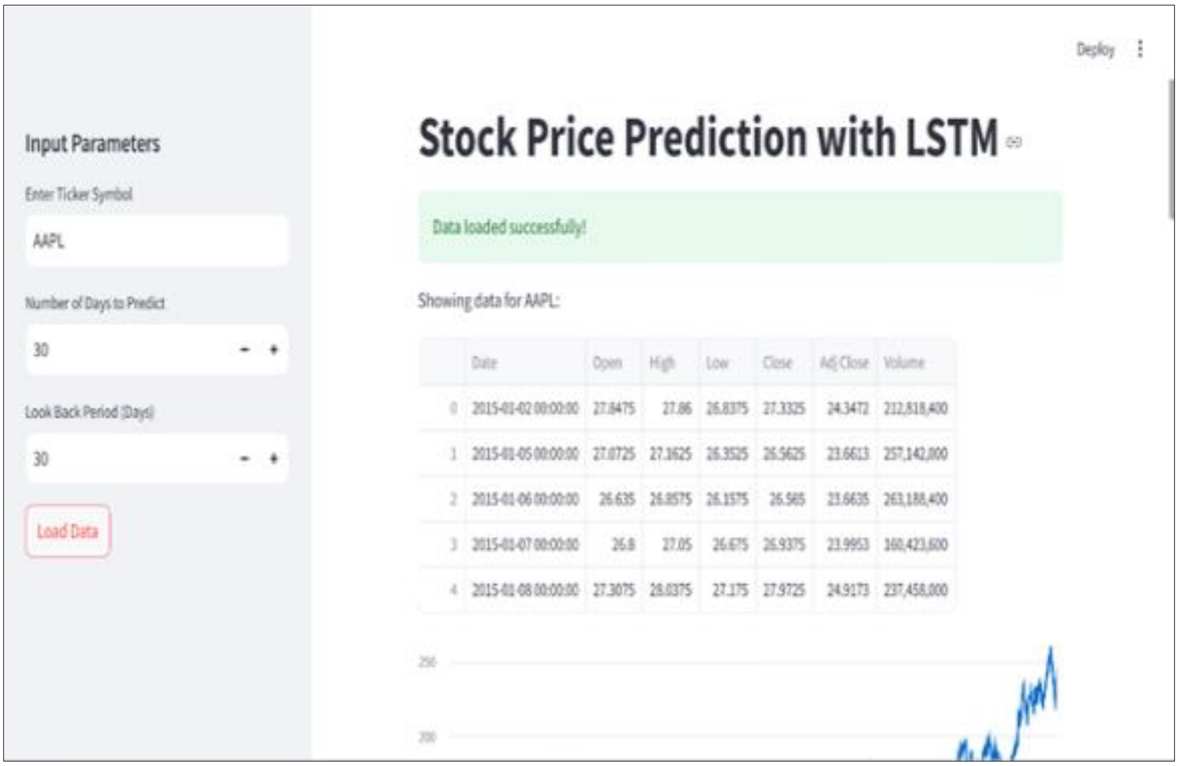

Figure 8 Stock price prediction description

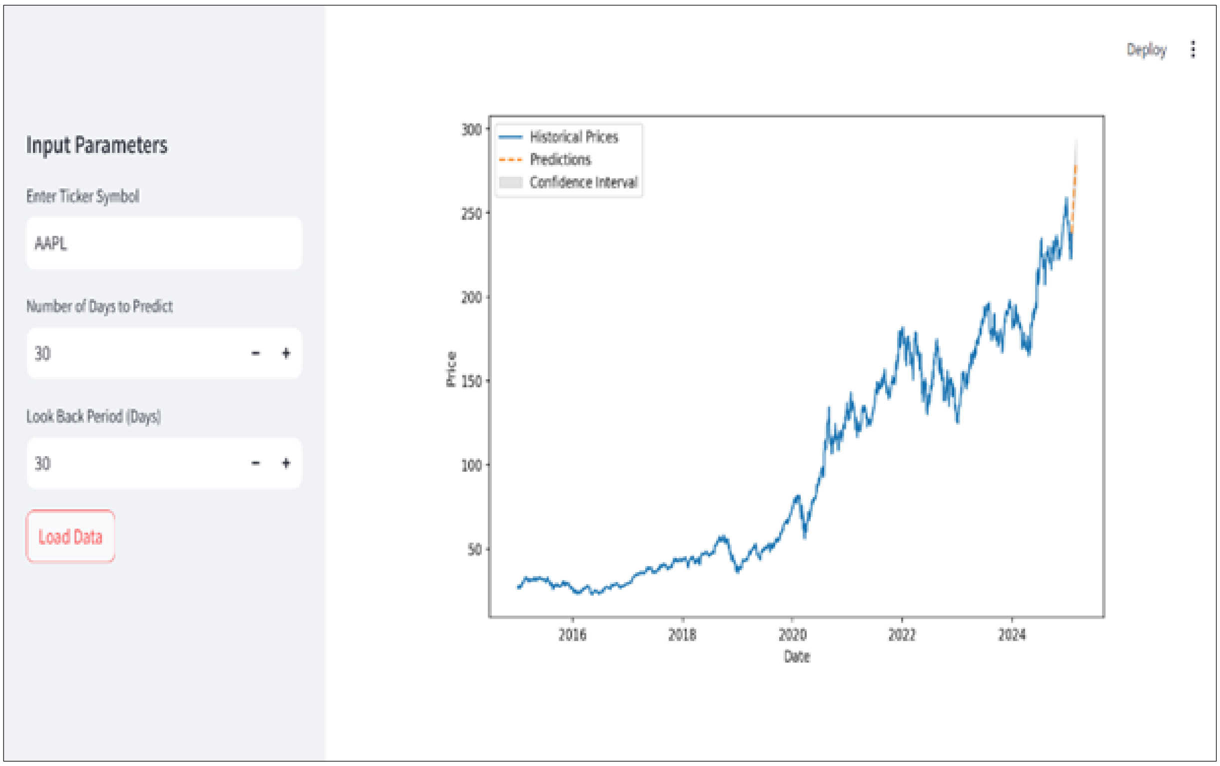

Figure 9 Final output of predicted stock price of 30 days

7. Conclusion

In conclusion, the Efficient Market Hypothesis (EMH) asserts that financial markets fully reflect all available information, making it impossible to consistently achieve above-average returns through prediction or active trading. Empirical studies and machine learning models, such as LSTMs, often struggle to generate reliable excess returns, supporting EMH. However, observed market anomalies, behavioral biases, and occasional predictive success challenge its strictest forms. While EMH remains a foundational theory in finance, real-world market inefficiencies suggest that some investors may exploit short-term opportunities, though doing so consistently remains difficult. Ultimately, while EMH provides a strong theoretical framework, real-world deviations highlight the complexity of financial markets and the potential for short-term inefficiencies.

Compliance with ethical standards

Acknowledgments

The author expresses sincere gratitude to Mrs. Niharika Pilligundla for their valuable contributions to data collection, model evaluation, and insightful discussions that helped shape this research.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

Statement of Ethical approval

This research was conducted in accordance with ethical guidelines to ensure transparency, integrity, and responsible handling of data. No sensitive personal data or human subjects were involved in this study.

References

[1] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

[2] Kim, J., Kim, J., & Kang, S. (2014). Stock price prediction using KNN and neural network. Journal of Advances in Information Technology, 5(3), 171-178..

[3] F Lai, C. C., & Yang, C. S. (2011). Stock price prediction using hybrid machine learning techniques. Expert Systems with Applications, 38(11), 14314-14323.

[4] Lee, T. S., & Chen, W. C. (2005). Mining the stock market: which measure is the best predictor of stock prices?. Expert Systems with Applications, 28(4), 743-754.

[5] Manavalan, E., & Sathiyamoorthy, E. (2012). Stock price prediction using K-Nearest Neighbor (k-NN) algorithm. International Journal of Computer Science and Engineering, 1(2), 88-93.

[6] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63..

[7] M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016