

# Stock Market Price Prediction Using NLP with Sentimental Analysis

Harshal Prashant Jangam<sup>1</sup>, Reshma Dilip Kuchekar<sup>2</sup>, Hrushikesh Nitin Kumthekar<sup>3</sup>

Shraddha Krishna Burde<sup>4</sup>, Prof. Mayuri S. Agrawal<sup>5</sup>

Students, Department of Computer Engineering<sup>1,2,3,4</sup>

Professor, Department of Computer Engineering<sup>5</sup>

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

**Abstract:** *Stock Many people use stock investments as a means of increasing their wealth. However, market downturns can cause huge losses and need to be predicted for a timely sell. In fact, with effective prediction, stocks are a good investment even during periods of market volatility as many stocks are —on sale. News is a significant signal source for changes in stock prices. However, stock analysts usually adjust their analysis according to the news in a subject manner, and wrong judgments can cause investors huge losses. Twitter is a terrific place to find breaking news and gets stock trend information in real time. Because there are so many Twitter users, news on the platform frequently has a significant impact on the market. In order to promptly incorporate Twitter news about a company into a time series prediction model on the company's stock price, this study suggests a data-driven pipeline. Our method, known as BERT-LSTM (BELT), uses the cutting-edge natural language processing (NLP) model BERT to extract useful features on stock price direction from Twitter news. These features are then used as covariates to a many-to-many stacked LSTM model that also uses historical stock prices to predict the direction of future stock prices. We adjust BERT using a carefully vetted stock news collection to. It is possible to efficiently identify the news tweets that are pertinent and to extract NLP elements that are predictive of price changes. In order to create a data-driven and objective pipeline to include news signals and prevent subjective analysis, all model parameters are trained from beginning to end. Extensive experiments on real stock prices and Twitter news show that BELT is able to predict stock prices more accurately utilizing news information than if historical price data are utilised alone for prediction, and outperforms StockNet, the state-of-the-art system for predicting news-based stock movement.*

**Keywords:** Stock Market

## I. INTRODUCTION

### 1.1 Overview

Stock market has a significant influence on both individual investor's interest and the country's entire economy trend. Stock movement prediction plays an important role in reducing risk of loss and increasing gains. Stock prediction is always regarded as a challenging work. Due to the high volatility and uncertainty of complex market dynamics, internal and external information changes constantly and it is hard to make the best prediction, the security current price reflects all market information and if there is a new change in financial market, the stock price will adjust immediately. Two approaches that works on stock market forecasting: fundamental and technical analysis.

### 1.2 Motivation

- We get motivate by the short comings of the current system.
- The goal of the system is to create something that can examine itself and make predictions based on News By Using NLP.
- The problem of stock price prediction is a well-known and essential one.

With a successful stock forecasting model, we can gain knowledge about market behavior overtime, identifying trends that might otherwise go unnoticed don't he basis of News Data

## II. STOCK MARKET PRICE PREDICTION

Stock market price prediction is a difficult task that involves analyzing past trends, financial data, and using algorithms to make predictions about future price movements. Different methods such as technical analysis, fundamental analysis, and machine learning can be used to predict stock prices. However, it's important to remember that no method can guarantee accurate predictions, and there is always a level of risk involved in investing in the stock market. Stock market price prediction using NLP (Natural Language Processing) with sentimental analysis is a method that involves analyzing textual data, such as news articles, social media posts, and financial reports, to predict stock prices. This method combines NLP techniques, which involve processing and understanding human language, with sentimental analysis, which involves analyzing the emotional tone of the text to identify positive or negative sentiment.

The key steps involved in stock market price prediction using NLP with sentimental analysis include:

1. Data collection: Collect textual data related to the stock, including news articles, social media posts, and financial reports.
2. Pre-processing: Pre-process the data by removing noise, stop words, and punctuation, and convert the text to a numerical format that can be analyzed by a machine learning algorithm.
3. Sentimental analysis: Use sentimental analysis techniques to determine the emotional tone of the text, such as positive, negative, or neutral sentiment.
4. Feature extraction: Extract features from the pre-processed text data that are relevant to the stock market, such as the frequency of certain words or phrases.
5. Model training: Train a machine learning algorithm, such as a neural network or support vector machine, to learn the relationship between the features and the stock price.
6. Prediction: Use the trained model to make predictions about future stock prices based on the pre-processed textual data and the sentiment analysis results.
7. Evaluation: Evaluate the performance of the model using metrics such as accuracy, precision, and recall.

## III. SUPPORT VECTOR MACHINE

A supervised machine learning approach called the Support Vector Machine (SVM) is useful for both classification and regression analysis. The goal of SVM is to find the optimal hyper plane that separates data points of different classes in a high-dimensional space. In classification, SVM can be used to predict the class of a new data point based on its position relative to the hyper plane. In regression analysis, SVM can be used to predict a continuous output variable based on the input variables.

Support Vector Machines (SVMs) can be used in Stock Market Price Prediction using NLP with Sentimental Analysis by classifying the text data as positive or negative sentiment and using it as a feature in the SVM model. The SVM model can then use the positive or negative sentiment feature along with other relevant features to predict the stock price.

In Stock Market Price Prediction using NLP with Sentimental Analysis, Support Vector Machines (SVMs) can be used to predict stock prices based on textual data. The first step is to collect relevant textual data, such as news articles, social media posts, and financial reports. The collected data is then pre-processed to remove stop words, punctuation, and special characters and convert the text to a numerical format that can be analyzed by the SVM model. Sentimental analysis techniques are then used to determine the emotional tone of the text, such as positive, negative, or neutral sentiment. Relevant features are extracted from the pre-processed textual data, such as the frequency of positive or negative sentiment words and other financial data features. An SVM model is then trained on the extracted features and the stock price data to learn the relationship between them. The trained SVM model is tested on the testing dataset to evaluate its performance. Finally, the hyper parameters of the SVM model are adjusted to improve its performance, such as the regularization parameter or kernel type. However, it is important to note that the accuracy of sentimental analysis and the quality and size of the training dataset can significantly affect the performance of the SVM model.

## IV. NATURAL LANGUAGE PROCESSING

NLP (Natural Language Processing) is a key component in Stock Market Price Prediction using NLP with Sentimental Analysis. NLP techniques are used to pre-process the textual data by converting the raw text into a numerical format

that can be analyzed by machine learning models. This involves tasks such as tokenization, where the text is split into individual words, removing stop words, punctuation, and special characters, and stemming, where words are reduced to their base form.

NLP techniques are also used for sentimental analysis, which determines the emotional tone of the text, such as positive, negative, or neutral sentiment. Sentimental analysis is important in stock market price prediction because news and other events can have a significant impact on stock prices. By incorporating sentiment as a feature in machine learning models, such as SVM, the models can potentially capture the impact of these events on stock prices.

**Algorithm for Natural Language Processing (NLP):**

1. Text preprocessing:
  - a. Tokenization: split the text into words or phrases.
  - b. Lowercasing: convert all text to lowercase.
  - c. Stopword removal: remove common words such as "the", "is", "a", etc.
  - d. Stemming/Lemmatization: reduce words to their base form (e.g. "running" -> "run").
2. Feature extraction:
  - a. Bag of Words: count the frequency of each word in the text.
  - b. TF-IDF (Term Frequency-Inverse Document Frequency): weight words by how frequently they appear in the text and how frequently they appear in the overall corpus.
3. Text classification:
  - a. Choose a classification algorithm, such as Naive Bayes, SVM, or Random Forest.
  - b. Train the algorithm on a labeled dataset.
  - c. Test the algorithm on a held-out test dataset to evaluate its performance.
4. Sentiment analysis:
  - a. Use a pre-trained sentiment analysis model or train a model on a labeled dataset.
  - b. Apply the model to the text to classify it as positive, negative, or neutral.
5. Named entity recognition:
  - a. Identify in the text any entities, including people, locations, and organizations.
  - b. Use a pre-trained model or train a model on a labeled dataset.
6. Text generation:
  - a. Use a pre-trained language model or train a model on a large text corpus.
  - b. Generate text by conditioning the model on a starting phrase or sentence.

**V. OUTPUTS**

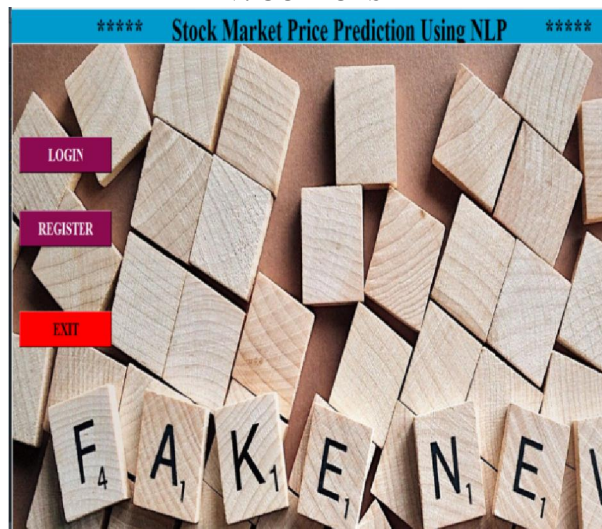


Fig 1. GUI Main Page  
DOI: 10.48175/568



**Registration Form**

Full Name :

Address :

E-mail :

Phone number :

Gender :  Male  Female

Age :

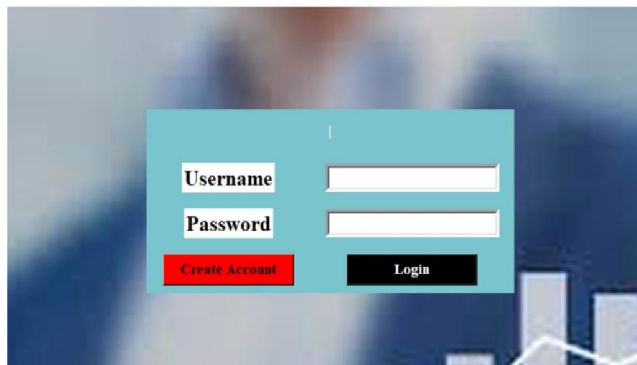
User Name :

Password :

Confirm Password :

**FAKE NEWS**

Fig 2. Registration Page



Username

Password

Fig3. Login Page

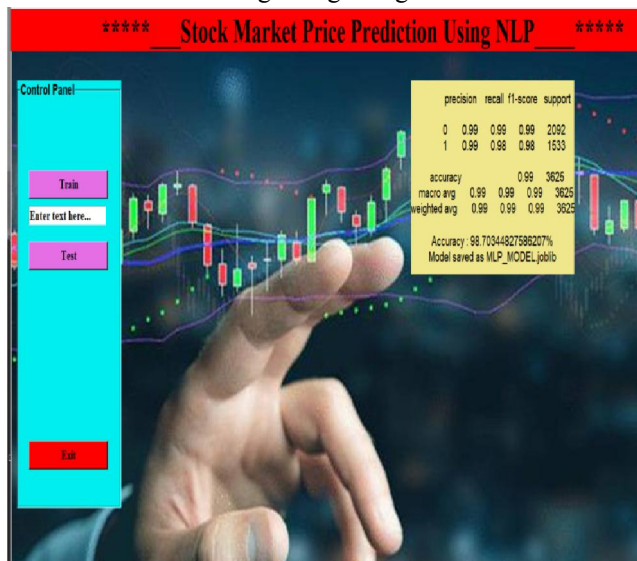


Fig4. Train Output

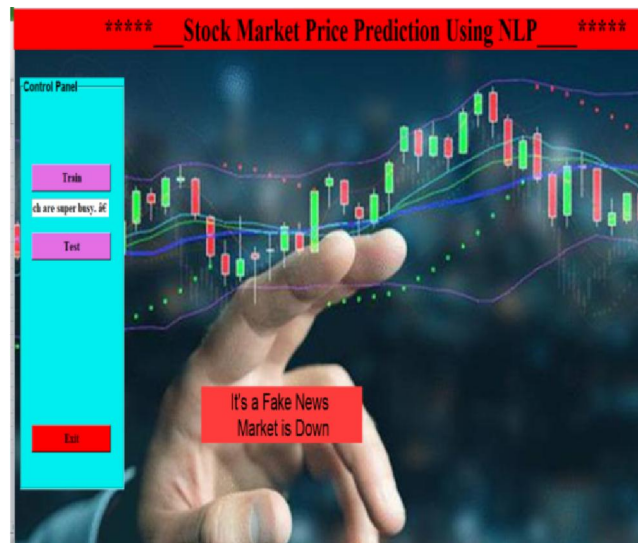


Fig5.1. Test Output

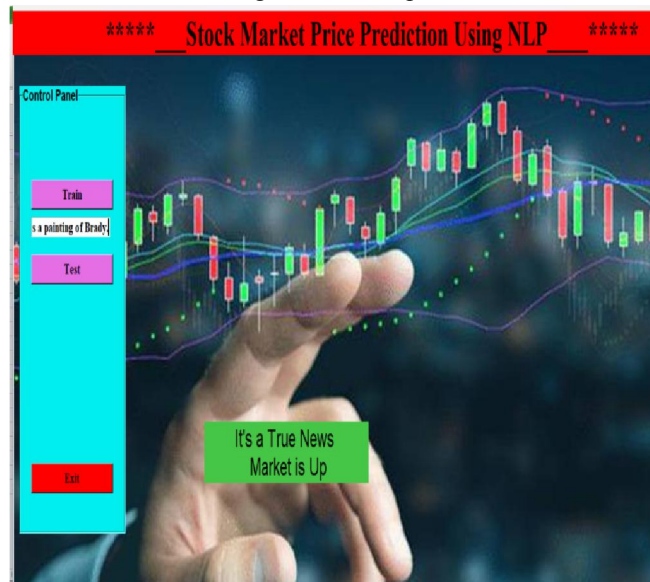


Fig 5.2 Test Output

## VI. CONCLUSION

In this project using augmented textual features as well as using finer- grained textual and sentiment analysis would provide better predictive ability to discover the stock market movement direction Based on News Headlines using Natural Language Processing. In the current work, we have proposed a methodology that starts by extracting multiple text-based features to enrich the representation of sentiments.

## VII. ACKNOWLEDGMENT

Express my true sense of gratefulness, sincere to my companion to the design Prof. Mrs. Mayuri Agrawal for his precious collaboration and guidance that she gave me during my exploration, to motivate me and give me access to all the laboratory coffers, which enabled me to conduct this exploration in a veritably straightforward and useful manner. I would also like to express my thanks to our fellow, Prof. R.H. Borhade HOD Sir, and Principle Dr. A.V. Deshpande and all of my musketeers who supported me during my laborious labour, whether they knew it or not.

**REFERENCES**

- [1]. P. D. Yoo, M. H. Kim, and T. Jan, —Machine learning techniques and use of event information for stock market prediction: A survey and evaluation,| in Proc. IEEE Int. Conf. Comput. Intell. Modelling, Control Automat., Int. Conf. Intell. Agents, Web Technol. Internet Commerce, vol. 2, Nov. 2005, pp. 835–841.
- [2]. Ravi and V. Ravi, —A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, | Knowl.-Based Syst., vol. 89, pp. 14–46, Nov. 2015.
- [3]. M. Zembik, —Social media as a source of knowledge for customers and enter- prises,| Online J. Appl. Knowl. Manage., vol. 2, no. 2, pp. 132–148, Aug. 2014.
- [4]. M. R. Vargas, C. E. M. dos Anjos, G. L. G. Bichara, and A. G. Evsukoff,—Deep learning for stock market prediction using technical indicators and fi- nancial news articles,| in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.
- [5]. Z. S. Bouktif and M. A. Awad, —Ant colony based approach to predict stock market movement from mood collected on Twitter,| in Proc. IEEE/ACM Int. Conf. Adv. S
- [6]. L. Bing, K. C. C. Chan, and C. Ou, —Public sentiment analysis in Twitter data for prediction of a Company’s stock price movements,| in Proc. IEEE 11th Int. Conf. E-Bus. Eng., Nov. 2014, pp. 232–239.
- [7]. Y. Saeys, I. Inza, and P. Larranaga, — A review of feature selection techniques in bioinformatics, | Bioinformatics, vol. 23, no. 19, pp. 2507–2517, Aug. 2007.
- [8]. J. V. Tu, —Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes,| J. Clin. Epi- demiol., vol. 49, no. 11, pp. 1225–1231, Nov. 1996.
- [9]. L. Oneto, Model Selection and Error Estimation in a Nutshell. Cham, Switzer- land: Springer, 2020.