IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stock Market Prediction Using Twitter Sentiment Analysis: A Review

Snehal Hemantkumar Chavan, 2Prof. Anil Gujar

Student

SPPU

Abstract

In financial engineering, stock market prediction is a crucial subject, particularly given the ongoing advancement of innovative methods and strategies in this area. In this study, we look into how sentiment exhibited in Twitter tweets affects the ability to predict stock prices. Twitter is the social networking site that gives everyone a free forum to publicly share their opinions. In specifically, we use the API to retrieve the company's live Twitter tweets. Every special character and stop word in the dataset is taken out. The Random Forest classifier is used to analyze sentiment in the filtered data. As a result, the tweets are categorized as neutral, negative, and positive. To get the outcome, the machine learning model receives the stock data and the tweet data as input. The stock market price is predicted using the ML classifier as a model. The real stock market value is contrasted with the predicted value that was obtained. Experiments employing real-time Twitter data and daily stock data from a number of businesses. The project's objective is to forecast the future price of a stock of interest by combining sentiment analysis of news headlines and Twitter tweets with previous stock data.

Keywords: Stock Market, Data Analysis, Machine Learning, Stock price Prediction, Social Media.

Introduction:

Stock market prediction is a crucial area of research in financial engineering, driven by the continuous development of new techniques and methodologies. Given the vast amount of data generated daily, particularly on social media platforms like Twitter, researchers have increasingly turned to these sources to enhance prediction accuracy. Twitter, with millions of tweets posted each day, serves as a rich resource for sentiment analysis related to various companies. Additionally, traditional media sources, such as newspaper headlines, offer valuable insights that can inform stock predictions.

To facilitate this process, live Twitter data can be extracted using the Twitter API and analyzed with machine learning classifiers. Among the various machine learning algorithms available, Random Forest and Naïve Bayes have emerged as key classifiers for training predictive models, offering improved accuracy in forecasting stock prices.

Moreover, incorporating advanced methods like Type 2 fuzzy logic enhances the robustness and flexibility of prediction models by applying both technical and fundamental indices as input variables. This multi-faceted approach aims to minimize errors and optimize prediction outcomes, making significant strides in stock market analysis.

Motivation:

- Investors look to profit from equity portfolios by purchasing and selling their holdings at the opportune moment of maximum or minimum potential profit.
- Unlike traditional, physical statistics, the future price of a stock can be difficult to forecast.
- Stock market is extremely variable and nonlinear.
- The rapid growth in economic globalization.
- International markets, and emerging information technology are all resulting in new ways to expand the reach of this system research.

Literature Survey

A Mittal and a. Goel.,This paper explores stock price forecasting using sentiment analysis from social media. By collecting past tweets and applying machine learning techniques like Naive Bayes and SVM, the system evaluates tweet sentiment and its correlation with stock price behavior. The goal is to predict future market movements and assess the potential of contrarian investing. Initial results show a promising connection between tweet sentiment and stock prices.

Agarwal, Apoorv, et al., We examine sentiment analysis on Twitter data. The contributions of this paper are: (1) We introduce POS-specific prior polarity features. (2) We explore the use of a tree kernel to obviate the need for tedious feature engineering. The new features (in conjunction with previously proposed features) and the tree kernel perform approximately at the same level, both outperforming the state-of-the-art baseline.

Heesoo Hwang and Jinsung Oh.,This paper explores fuzzy models for predicting daily and weekly stock prices, addressing challenges in traditional time series analysis. Three fuzzy rule formats were developed using trapezoidal membership functions, with traditional candlestick-chart indicators as input variables. Optimal models were identified through differential evolution (DE) and tested on the Korea Composite

Stock Price Index (KOSPI), demonstrating promising predictive performance for open, high, low, and close prices.

N. Chethan and R. Sangeetha.,This paper uses R programming to perform sentiment analysis on tweets about USD/INR, BSE Sensex, and NSE Nifty. Tweets are classified into eight sentiment categories, and daily sentiment is identified as positive, negative, or neutral. Word clouds highlight key discussions, helping investors predict stock price movements based on market sentiment.

Troy J. Strader John J. Rozycki Thomas H. Root.,This study reviews stock market prediction using machine learning, categorizing research into four areas: neural networks, support vector machines, genetic algorithms, and hybrid approaches. Common findings and limitations are analyzed, with suggestions for future research directions in improving market forecasting.

FRÉDÉRIC C. G. BOGAERTS 1, NAGHMEH IVAKI 1 (Member, IEEE), AND JOSÉ FONSECA., A comprehensive dataset of 1,026 publicly disclosed Python vulnerabilities has been compiled to enhance security awareness and practices. These vulnerabilities are classified using frameworks like CWE and OWASP Top 10, offering a structured understanding of common flaws and risks. The dataset includes patched and vulnerable code samples, aiding developers, researchers, and security teams in identifying trends, addressing common errors, and mitigating threats. By fostering better vulnerability analysis and prevention, this resource aims to support the creation of more secure Python applications.

DINIS BARROQUEIRO CRUZ , JOÃO RAFAEL ALMEIDA , AND JOSÉ LUÍS OLIVEIRA., As software grows more complex and vulnerable to cyber threats, this work explores three key security approaches—SAST, DAST, and SCA—focusing on open-source solutions. It proposes a baseline comparison model and workflow for vulnerability assessments, highlighting challenges and opportunities to enhance application security against emerging threats.

Oshando Johnson, Eric Bodden., Dev-Assist, an IntelliJ IDEA plugin, leverages multi-label machine learning to detect security-relevant methods more accurately by considering label dependencies. It automates static analysis tool configuration, reducing manual effort and improving detection performance with a higher F1-Measure than existing methods.

NISREAN THALJI 1 , ALI RAZA 2 , MOHAMMAD SHARIFUL ISLAM 3 , NAGWAN ABDEL SAMEE 4 , AND MONA M. JAMJOOM., This study introduces AE-Net, an AI-driven approach using autoencoders for feature extraction to detect SQL injection attacks. With a 46,392-query dataset, the method achieves a 0.99 k-fold accuracy using extreme gradient boosting, validated through hyperparameter tuning and statistical analysis, offering a robust solution for automated SQL injection detection.

ONUR AKTAS 1 AND AHMET BURAK CAN., This study introduces a machine learning-based approach to improve web crawler efficiency by predicting the need for JavaScript rendering. Using a dataset from 17,160 websites, the method reduces execution time by 20% without compromising coverage, enhancing the effectiveness of security-focused web crawlers while optimizing resource usage.

Sr.No	Year	Author	Gap Analysis
1.	2024	FRÉDÉRIC C. G. BOGAERTS 1, NAGHMEH IVAKI 1 (Member, IEEE), AND JOSÉ FONSECA	Existing Python vulnerability data is fragmented, incomplete, and often lacks actionable insights, making it challenging for developers and security teams to address risks effectively. A unified, well-classified dataset bridges this gap, enabling better vulnerability detection, analysis, and mitigation.
2.	2023	DINIS BARROQUEIRO CRUZ , JOÃO RAFAEL ALMEIDA , AND JOSÉ LUÍS OLIVEIRA	Current application security tools often lack integration and comprehensive coverage, leaving gaps in vulnerability detection. This highlights the need for a unified approach combining SAST, DAST, and SCA to address emerging threats effectively.
3.	2024	Oshando Johnson, Eric Bodden	Current static analysis tools require tedious, error-prone manual configuration and overlook dependencies in security-relevant methods, limiting accuracy. Dev-Assist addresses these gaps with automated configurations and a dependency-aware machine learning approach.
4.	2023	NISREAN THALJI 1 , ALI RAZA 2 , MOHAMMAD SHARIFUL ISLAM 3 ,	Current SQL injection detection methods often lack automation and high accuracy. This study addresses these gaps with AE-Net, leveraging AI for feature extraction and

		NAGWAN ABDEL SAMEE 4 , AND MONA M. JAMJOOM.,	achieving superior detection performance.
5.	2024	ONUR AKTAS 1 AND AHMET BURAK CAN	Traditional web crawlers struggle with resource inefficiency and handling dynamic JavaScript content. This study addresses these gaps by predicting JavaScript rendering needs, improving crawler efficiency and reducing computational requirements.

Gap Analysis

- Data Quality: Twitter data can be noisy and may contain irrelevant or misleading information that can affect prediction accuracy.
- Sentiment Analysis Accuracy: The effectiveness of sentiment analysis depends on the classifier's ability to accurately interpret the emotional tone of tweets, which can be subjective.
- Market Influences: External factors such as political events, economic indicators, and sudden market shifts may not be captured by social media or historical data.
- Computational Complexity: Advanced machine learning models like XGBoost can be resource-intensive, requiring significant computational power and time for training.
- Limited Historical Data: The availability of historical stock prices and corresponding social media data may be limited, impacting model training.

Proposed System

The high-level framework of the stock market prediction utilizing sentiment analysis from Twitter is shown in Figure 1. The option to obtain the anticipated stock price of the relevant company on the stock market is provided to the user. The name of the company whose stock price has to be predicted must be entered by the user. Additionally, the user can observe the market's active stocks as well as the weekly stock market analysis. Two primary datasets are used in this investigation. Although we are gathering data for accuracy purposes, we are retrieving data from Twitter. These two datasets are what we have collected. We have eliminated any special characters, such as emoticons, hashtags (#), and @, from this data. These special characters are eliminated because they are not required for sentiment, and we have only looked at simple sentences. The tweets are divided into three categories when sentiment analysis is done in machine learning: neutral, negative, and positive. Line by line, it retrieves the lexical file data, and it will also retrieve the data from Twitter and newspaper headlines that we are obtaining.

Proposed System Diagram

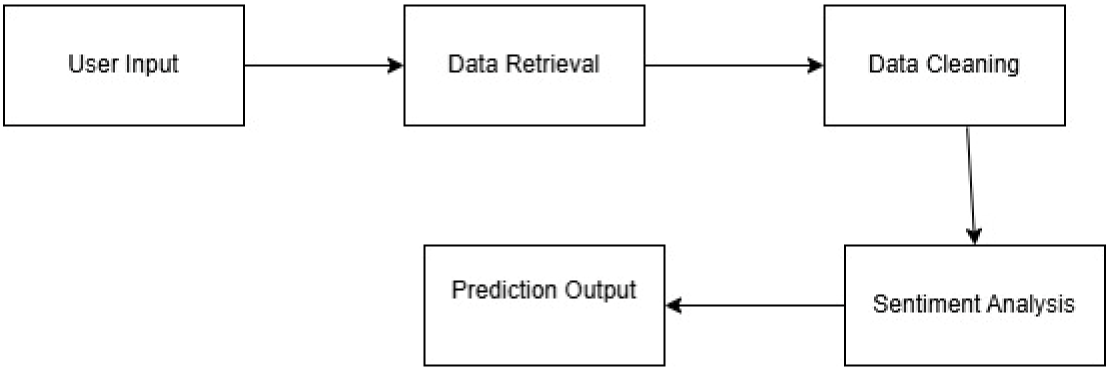

Fig: Stock market prediction system.

Conclusion

In this paper we investigated how sentiment analysis of the twitter data is correlated to the prediction of the stock market price for all the companies which are taken. The result obtained after the prediction process clearly specifies that, we have obtained the accurate value which matches with the actual stock price appropriately.

References:

[1]B.Weng, M. A. Ahmed, and F. M. Megahed, ‘‘Stock market one-day ahead movement prediction using disparate data sources,’’ Expert Syst. Appl., vol. 79, pp. 153–163, Aug. 2017.

[2] M. S. Checkley, D. A. Higón, and H. Alles, ‘‘The hasty wisdom of the mob: How market sentiment predicts stock market behavior,’’ Expert Syst. Appl., vol. 77, pp. 256–263, Jul. 2017.

[3] T. M. Nisar and M. Yeung, ‘‘Twitter as a tool for forecasting stock market movements: A short-window event study,’’ J. Finance Data Sci., vol. 4, no. 2, pp. 101–119, Jun. 2018.

[4] S. Erfanian, Y. Zhou, A. Razzaq, A. Abbas, A. A. Safeer, and T. Li, ‘‘Predicting Bitcoin (BTC) price in the context of economic theories: A machine learning approach,’’ Entropy, vol. 24, no. 10, p. 1487, Oct. 2022.

[5] X. Zhong and D. Enke, ‘‘A comprehensive cluster and classification mining procedure for daily stock market return forecasting,’’ Neurocomputing, vol. 267, pp. 152–168, Dec. 2017.

[6] FRÉDÉRIC C. G. BOGAERTS 1, NAGHMEH IVAKI 1 (Member, IEEE), AND JOSÉ FONSECA, ‘‘A Taxonomy for Python Vulnerabilities,’’ 2024, arXiv:1911.09359.

[7] DINIS BARROQUEIRO CRUZ , JOÃO RAFAEL ALMEIDA , AND JOSÉ LUÍS OLIVEIRA, ‘‘Open Source Solutions for Vulnerability Assessment: A Comparative Analysis’’ in Proc. 5th Int. Conf. Mechatronics Comput. Technol. Eng. (MCTE), Dec. 2023, pp. 1173–1178.

[8] Oshando Johnson, ‘‘Detecting Security-Relevant Methods using Multi-label Machine Learning’’ Expert Syst. Appl., vol. 157, Nov. 2024, Art. no. 113481

[9] NISREAN THALJI 1 , ALI RAZA 2 , MOHAMMAD SHARIFUL ISLAM 3 , NAGWAN ABDEL SAMEE 4 , AND MONA M. JAMJOOM.,” AE-Net: Novel Autoencoder-Based Deep Features for SQL Injection Attack Detection”.,2023.

[10] ONUR AKTAS 1 AND AHMET BURAK CAN.,” Making JavaScript Render Decisions to Optimize Security-Oriented Crawler Process”.,2024.