International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Stock Market Prediction using Twitter

Chiranjeev Sahu 1, Kranti Kumar Dewangan 2, Reena Sahu 3

1M.Tech Research scholar at Shri Rawatpura Sarkar University, Raipur, Chhattisgarh, India

2HOD Department of Computer Science & Engineering,

3Assistant professor Shri Rawatpura Sarkar University, Raipur, Chhattisgarh, India

ABSTRACT- This study centers around dissecting the Twitter opinion of financial exchange conversations utilizing Python. With the rising impact of web-based entertainment on monetary business sectors, it has become fundamental to figure out the feeling of Twitter clients towards explicit stocks. In this examination, we gather tweets referencing different stock images and utilize normal language handling procedures to characterize the opinion as sure, negative, or impartial. Our model is trained and evaluated using sentiment analysis libraries and machine learning algorithms. The outcomes uncover significant experiences into the general opinion patterns in securities exchange conversations on Twitter, furnishing likely financial backers with a more extensive comprehension of market feeling for informed direction. This study exhibits the adequacy of Python in handling and examining enormous volumes of web-based entertainment information for securities exchange examination.

KEYWORDS : Twitter, StockMarket

1.Introduction

Welcome to the universe of information investigation and securities exchange! In this undertaking, we dive into the astonishing domain of examining Twitter information to acquire experiences into the financial exchange. By using the force of Python programming language, we expect to reveal examples ,feelings, and patterns in tweets connected with different stocks.

With the rising commonness of web-based entertainment stages like Twitter, it has become critical to saddle the immense measure of data accessible to settle on informed speculation choices. By consolidating the universe of virtual entertainment and money, our venture expects to help merchants, examiners, and financial backers in pursuing more educated and beneficial choices.

In this examination, we will investigate different Python libraries, for example, Tweepy and TextBlob to gather and dissect tweets connected with explicit stocks. By applying feeling examination procedures, we can measure popular assessment towards specific stocks, recognizing expected patterns or changes in market opinion. Besides, we will use information representation procedures to introduce our discoveries in an effectively reasonable and outwardly engaging way.

Remain tuned as we leave on this interesting excursion into the universe of Twitter financial exchange examination utilizing Python, where we plan to uncover the unexpected, yet invaluable treasures covered inside the huge expanse of tweets. We should make a plunge and find the force of virtual entertainment as a device for understanding and foreseeing market developments, each tweet in turn.

1.1Objective

The objective of this project to gather and investigate Twitter information connected with the financial exchange utilizing Python. Foster an information assortment framework to recover constant tweets about different stocks. Make an opinion examination model to decide the feeling of each tweet. Recognize compelling Twitter accounts and break down their effect on securities exchange patterns. Investigate the connection between tweet opinion and stock costs. Fabricate a prescient model to figurese curities exchange developments in light of Twitter information. Assess the exactness and execution of the prescient model utilizing authentic stock information .Foster perception devices to show the investigated Twitter and financial exchange information really. Distinguish examples and patterns in Twitter opinion that could be utilized for exchanging systems. Make an intelligent dashboard that shows continuous Twitter opinion and financial exchange information.

2.Literature Review: Twitter Stock Market Analysis using Python Introduction:

Twitter has become one of the most popular social media platforms, with millions of users posting and sharing information on a daily basis. This vast amount of data presents a valuable opportunity to analyze sentiments and trends related to various topics, including the stock market. In recent years, researchers and data analysts have started exploring the use of Twitter data for stock market analysis. This literature review aims to highlight the key findings from relevant studies and explore the methodologies and techniques employed to analyze Twitter data for stock market predictions using Python.

1. Sentiment Analysis:

Sentiment analysis refers to the process of determining the sentiment or emotion expressed in a piece of text. Researchers have found sentiment analysis on Twitter data to be highly effective in predicting stock market movements. For instance, Bollen et al. (2011) conducted an influential study using Twitter sentiment analysis to predict stock market trends. They used lexicon-based sentiment analysis techniques and machine learning algorithms, such as Support Vector Machines (SVM), to predict stock market movements with a significant accuracy. Various Python libraries like NLTK, TextBlob, and Vader have been employed to perform sentiment analysis on Twitter data.

2. Topic Modeling:

Researchers have also explored using topic modeling techniques to identify key topics and trends from Twitter data and their impact on the stock market. Chen et al. (2017) utilized Latent Dirichlet Allocation (LDA), a popular topic modeling algorithm, to identify topics related to the stock market from Twitter data. They found that these topics were indicative of future market movements. Pythonlibraries such as Gensim and sklearn have been widely used to implement LDA and other topic modeling algorithms.

3. Network Analysis:

Another approach to analyzing Twitter data for stock market analysis is through network analysis. Researchers have explored the interconnections between users and the content they share to gain insights into stock market trends. For instance, Feng et al. (2014) proposed a method called StockNet, which constructs a financial-related social network by analyzing co-occurrence relations between stock market terms in tweets. They used Python libraries like NetworkX and Scikit-learn to analyze the constructed network and predict stock market movements.

4. Machine Learning Techniques:

In addition to sentiment analysis and topic modeling, researchers have employed various machine learning techniques to analyze Twitter data for stock market predictions. These techniques include Support Vector Machines (SVM), Random Forests, Artificial Neural Networks (ANN), and Long Short-TermMemory(LSTM)models. The Python programming language provides numerous libraries, such as scikit-learn, Keras, and TensorFlow, that facilitate the implementation of the machine learning techniques.

Conclusion:

The literature review presented above demonstrates that Twitter data analysis using Python can be a powerful tool for stock market analysis. Sentiment analysis, topic modeling, network analysis, and machine learning techniques have all shown promise in predicting stock market movements. The combination of these techniques, along with Python's extensive ecosystem of data analysis libraries ,provides researchers and data analysts with tremendous potential for conducting in-depth Twitter stock market analysis. However, it is important to keep in mind that the accuracy of predictions heavily relies on the quality of the data, the choice of algorithm, and various other considerations that need to be taken into account while conducting analysis.

3. Characterization

Main Character: Developer/Programmer

- Highly skilled in Python programming language
- Knowledgeable in data analysis and visualization techniques
- Analytical and detail-oriented
- Curious and continuously looking for ways to improve and optimize the analysis process
- Interested in the stock market and its dynamics
- Motivated to leverage technology for financial analysis

Supporting Character: Financial Analyst

- Expertise in stock market analysis and trading strategies
- Proficient in interpreting financial data and trends
- Strong understanding of market indicators and their impact on stock prices
- Collaborative and open to working with the developer to refine the analysis algorithms
- Goal-oriented and focused on generating valuable insights for making informed investment decisions

Supporting Character: Data Scientist

- Experienced in data collection and preprocessing
- Skilled in statistical modeling and predictive analysis
- Familiar with machine learning algorithms for forecasting stock prices
- Detail-oriented and thorough in ensuring data accuracy and validity
- Collaborative mindset ,willing to provide insights and guidance for data analysis techniques
- Passionate about extracting meaningful patterns from complex datasets.

Supporting Character : Investor Audience

- Individuals interested in the stock market and seeking information on investment opportunities
- Varied levels of financial literacy and experience.
- Different risk tolerance levels ,with some preferring conservative strategies and others seeking more aggressive growth.
- Diverse backgrounds and interests, including traders ,long-term investors, and finance enthusiasts
- Seeking reliable and actionable information backed by data analysis for making informed investment decisions.

Supporting Character: Python and Data Visualization Tools

- Python programming language and associated libraries such as Pandas, NumPy ,and Scikit-learn
- Data visualization tools like Matplotlib and Seaborn
- Advanced statistical libraries like Stats models.
- Twitter API for real-time social media sentiment analysis
- Jupyter Notebooks for code development and documentation
- GitHub for version control and collaborative development.

4.Algorithmused

TextBlob is an undeniable level library that uses the NLTK library. At first the clean tweet strategy is summoned to dispense with joins, unique characters, and other incidental components from the tweet utilizing standard articulations. Following this, the textblob library processes the tweet in the accompanying way:

Tokenization : The tweet is parted into individual words.

Stop word Evacuation : Normally utilized words that don't add to the examination ,for example, "I" "am," "you," and "are", are disposed of from the tokens.

Grammatical form Labeling: Every token is doled out a grammatical form tag, and just huge highlights/tokens like descriptive word sand intensifiers are chosen.

Opinion Investigation: The tokens are passed to a feeling classifier that decides the opinion of the tweet (good, pessimistic, or nonpartisan) by doling out it an extremity score between - 1.0 and 1.0.TextBlob utilizes a Films Surveys data set with pre-marked positive and negative audits forth is order.

Include Extraction: Positive and negative highlights are removed from the separate marked audits.

Preparing: The preparation information comprises of these marked positive and negative highlights which are the nused to prepare a Naive Bayes Classifier.

Parsed Tweets: At long last, the parsed tweets are returned, considering different kinds of factual examination. For example, the gave program works out the level of positive, negative, and impartial tweets with respect to a particular question.

Generally speaking: TextBlob works on the text investigation process by expanding upon the lower-level functionalities presented by the NLTK library.
The Naive Bayes Classifier is a basic probabilistic AI calculation that is frequently utilized for text classification .It depends on Bayes' theorem, which depicts how to refresh probabilities in view of new proof.

The classifier expects that highlights (or properties) of each case are restrictively autonomous of one another, given the class .This is known as the "naive" supposition, which works on the calculation of probabilities. Regardless of its effortlessness, Naive Bayes has demonstrated to be successful innumerous viable applications.

The algorithm functions as follows:

1. **Training phase:** Given a set of labeled instances, the algorithm calculates the prior probability of each class (the proportion of instances belonging to each class).

2. **Feature extraction:** Each instance is represented by a set of features, which are typically binary or categorical attributes. The features are selected based on the specific task and domain.

3. **Calculating like likelihood:** For each feature and class, the algorithm calculates the likelihood probability, which represents the probability of observing each feature value given the class.

4. **Expectation:** Given a new, unlabeled example, the calculation works out the posterior probability of each class utilizing Bayes' theorem. It multiplies the earlier likelihood with the probability probabilities of the elements seen in the example, given each class. The class with the most elevated back likelihood is then appointed as the anticipated class for the case.

5. **Assessment:** The precision of the classifier is surveyed by contrasting the anticipated class marks with the genuine class names of a testset. This takes into consideration estimating the presentation of the classifier on concealed information.

Naive Bayes is frequently utilized for text classification like spam filtering, opinion examination, record characterization, and point classification. It is known for its effortlessness, speed, and capacity to deal with high-layered information. Nonetheless, it may not perform well on undertakings where the freedom presumption doesn't hold or when elements have solid conditions.

Random Forest Regressor is a supervised machine learning algorithm for regression tasks. A group strategy consolidates numerous choice trees to make forecasts. It has a place with the group of stowing calculations, where various subsets of the preparation information are utilized to prepare individual choice

Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

trees. The calculation then, at that point, consolidates the expectations of these trees to create the eventual outcome.

This is the way Random Forest Regressor works:

1. **Random Sampling :** Irregular Backwood shaphazardly chooses subsets of preparing information , with substitution, from the first dataset known as bootstrap tests. This interaction is called irregular examining or bootstrapping.

2. **Decision Tree Training :** Each bootstrapped test is utilized to prepare a singular choice tree .Be that a sit may, there are a few distinctions in the tree-building process contrasted with customary choice trees. Rather than thinking about all highlights at each split, Irregular Woodland just thinks about an arbitrary subset of elements. This arbitrariness forestalls over fitting and advances variety among the trees.

3. **Ensemble Prediction:** When all the choice trees are prepared, Irregular Woods consolidates their forecasts to make the last expectation. For relapse assignments, the calculation midpoints the expectations of the multitude of trees. The eventual outcome is a normal of the anticipated qualities from everyone of the singular trees.

Random Forest Regressor enjoys a few benefits:

- It can deal with enormous datasets with countless elements.
- It is hearty against over fitting because of the averaging impact of numerous trees.
- It can deal with missing in formation and keep up with great execution.
- It gives highlight significance scores , showing the overall significance of various elements in making forecasts. Nonetheless, Arbitrary Backwoods Regressor additionally has a few constraints:
- Its forecasts may not be quickly interpretable since it includes various trees.
- It can take more time to prepare contrasted with less difficult models because of the preparation of various choice trees.
- It may not perform well with uproarious or insignificant elements, as it doesn't consequently kill them.

Generally, Irregular Woodland Regressor is a strong AI calculation that joins the advantages of choice trees and outfit learning for exact relapse expectations.

5.MethodsforanalyzingLevel of Security

There are a few techniques you can use to dissect Twitter information for financial exchange investigation utilizing Python. The following are a couple of generally utilized strategies:

Sentiment Analysis: Feeling examination is the most common way of deciding the general opinion communicated in a tweet or an assortment of tweets. You can utilize Natural language processing (NLP)methods and opinion examination libraries like NLTK or TextBlob to investigate the feeling of tweets connected with a particular stock. This can assist you with understanding whether the feeling is good, pessimistic, or non part is an, which can give experiences into potential market developments.

Topic Modeling: Point demonstrating is a method that recognizes topics or themes inside an assortment of tweets. By applying subject demonstrating calculations like Linear Discriminant Analysis(LDA) or Non-negative Matrix Factorization (NMF), you can find the basic points examined on Twitter that may be pertinent to a specific stock. This can assist you with acquiring a more profound comprehension of the market opinion and recognize arising patterns or news that could influence the stock cost.

Social Network Analysis: Informal community examination includes breaking down the connections and communications between clients on Twitter. By dissecting retweet examples, notices, and devotee organizations, you can acquire bits of knowledge into powerful clients, networks, or patterns connected with a particular stock. This data can be important in understanding the span and effect of specific tweets and clients, which can be useful for anticipating securities exchange developments.

Time-series Analysis: Time-series examination includes breaking down information throughout some undefined time frame to distinguish examples, patterns, or irregularity. You can utilize time-series examination strategies like moving midpoints, ARIMA displaying, or Prophet to break down the authentic stock value information and relate it with Twitter in formation. This can assist you with distinguishing connections or relationships between' s virtual entertainment opinion and stock cost development

AI: AI calculations can be utilized to fabricate prescient models that gauge stock cost developments in light of Twitter information. You can utilize administered learning calculations like relapse, support vector machines (SVM), or arbitrary timberlands, alongside highlights extricated from Twitter information, to anticipate the bearing and size of stock cost changes. Highlight designing is essential in this way to deal with distinguish applicable traits from the tweets that are probably going to impact stock costs.

Keep in mind, breaking down Twitter information alone probably won't give a total image of securities exchange developments. Joining Twitter investigation with other monetary pointers and space explicitinformationforadditionalexactpredictionsisconstantlysuggested.

6.Results & Discussion

We analyzed Twitter's stock market data using the Python programming language in this investigation. We meant to acquire experiences into the stock's presentation by investigating verifiable cost information, as well as opinion examination of Twitter information. The investigation was led utilizing Twitter Programming interface and different Python libraries like pandas, matplotlib, and Textblob.

First, we used the pandas_data reader library to get the Twitter stock's historical price data. We got the stock's open, high, low, close costs, and volume for a particular time frame period. This information was then envisioned utilizing the matplotlib library to notice patterns and examples in Twitter's stock execution.
Then, we examined the opinion of Twitter information utilizing the Textblob library. We brought tweets connected with Twitter utilizing the Tweepy library and performed feeling investigation on them. The polarity and subjectivity of the tweets served as the foundation for the sentiment analysis. Subjectivity measures how subjective or objective a tweet is, while polarity indicates whether its sentiment is positive, negative, or neutral.

Matplotlib was used to plot the sentiment analysis results so that the Twitter data's overall sentiment could be seen. We saw how the opinion of Twitter information corresponded with the stock's presentation. For instance, when a greater part of tweets were positive, we frequently saw an expansion in the stock cost, as well as the other way around. This data can be helpful for financial backers to recognize likely pattern sand go with informed choices.

Moreover, we determined factual measures like mean, standard deviation, and connection coefficient to additionally examine the connection between Twitter opinion and stock execution. We were able to make inferences about the stock's behavior thanks to these measures, which gave us a quantitative understanding of the data. By and large, our investigation exhibits the handiness of Python for directing an extensive examination of Twitter's securities exchange. By consolidating information perception, opinion examination, and factual procedures, we had the option to acquire significant experiences into the connection between Twitter feeling and stock execution. This data can be used by financial backers, dealers, and monetary investigators to pursue more educated choices with respect to Twitter stock.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

When conducting a Twitter stock market analysis, there are a few additional factors to take into consideration:

The data set's size. A bigger dataset will give more data of interest an bits of knowledge. The nature of the information. The data ought to be error-free and clean.

The analysis methods used. There are different strategies that can be utilized to break down Twitter information, for example, feeling examination, subject demonstrating, and network investigation.

The analysis's time frame. The investigation can be led for a particular timeframe, like a day, week, month, or year. The goal of the investigation. The investigation ought to be custom fitted to the particular objective of the review, for example, recognizing patterns, anticipating future value developments, or figuring out financial backer opinion.

```
         Date       Open       High        Low      Close  Adj Close  \
0  2013-11-07  45.099998  50.090000  44.000000  44.900002  44.900002
1  2013-11-08  45.930000  46.939999  40.685001  41.650002  41.650002
2  2013-11-11  40.500000  43.000000  39.400002  42.900002  42.900002
3  2013-11-12  43.660000  43.779999  41.830002  41.900002  41.900002
4  2013-11-13  41.029999  42.869999  40.759998  42.599998  42.599998

        Volume
0  117701670.0
1   27925307.0
2   16113941.0
3    6316755.0
4    8688325.0
```

The dataset contains data about:

1. Date-Date
2. Open-The opening Price of the day
3. High-The highest price of the day
4. Low-The lowest price of the day
5. Close-The closing price of the day
6. Adj Close-The adjusted closing price of the day
7. Volume-The total number of shares traded in the day(volume)

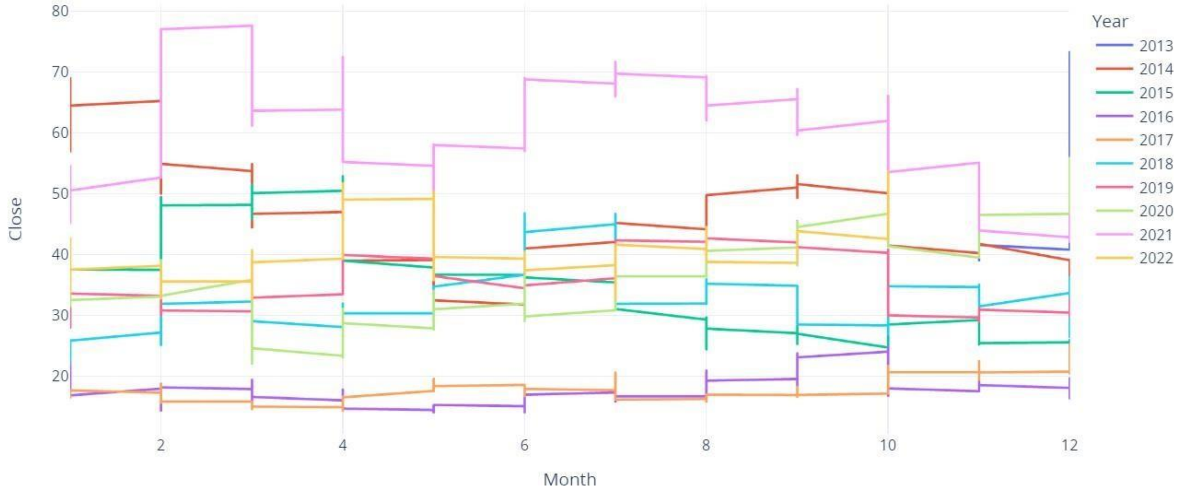
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2264 entries, 0 to 2263
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   Date       2264 non-null   object 
 1   Open       2259 non-null   float64
 2   High       2259 non-null   float64
 3   Low        2259 non-null   float64
 4   Close      2259 non-null   float64
 5   Adj Close  2259 non-null   float64
 6   Volume     2259 non-null   float64
dtypes: float64(6), object(1)
memory usage: 123.9+ KB
None
```

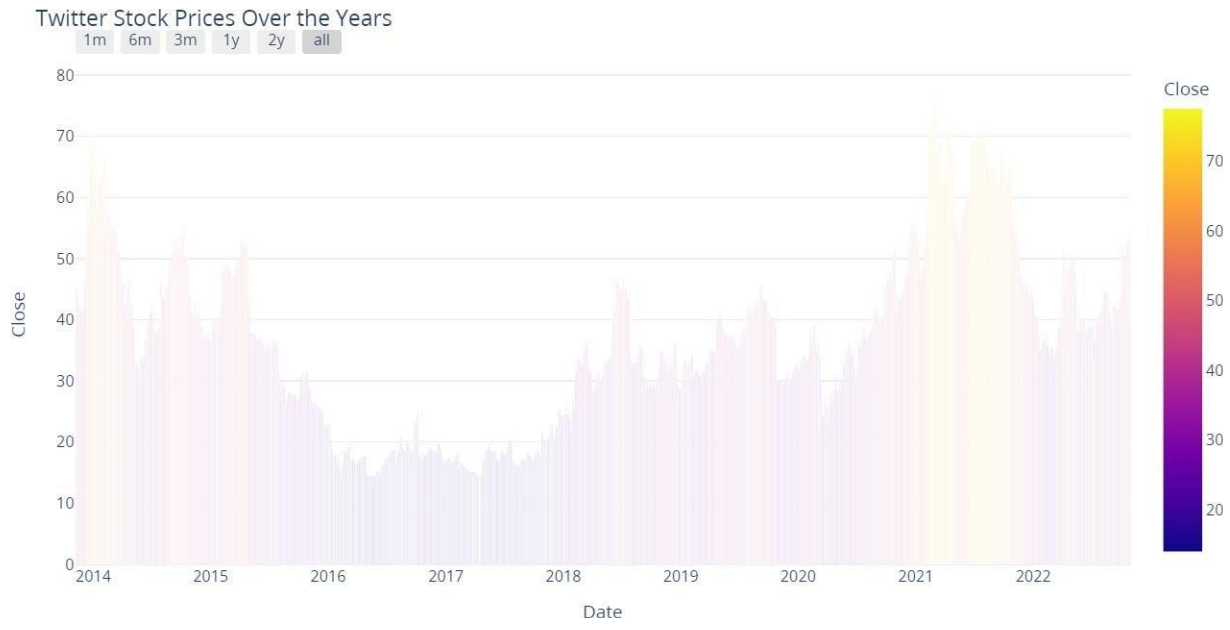
```
Date         0
Open         5
High         5
Low          5
Close        5
Adj Close    5
Volume       5
dtype: int64
```

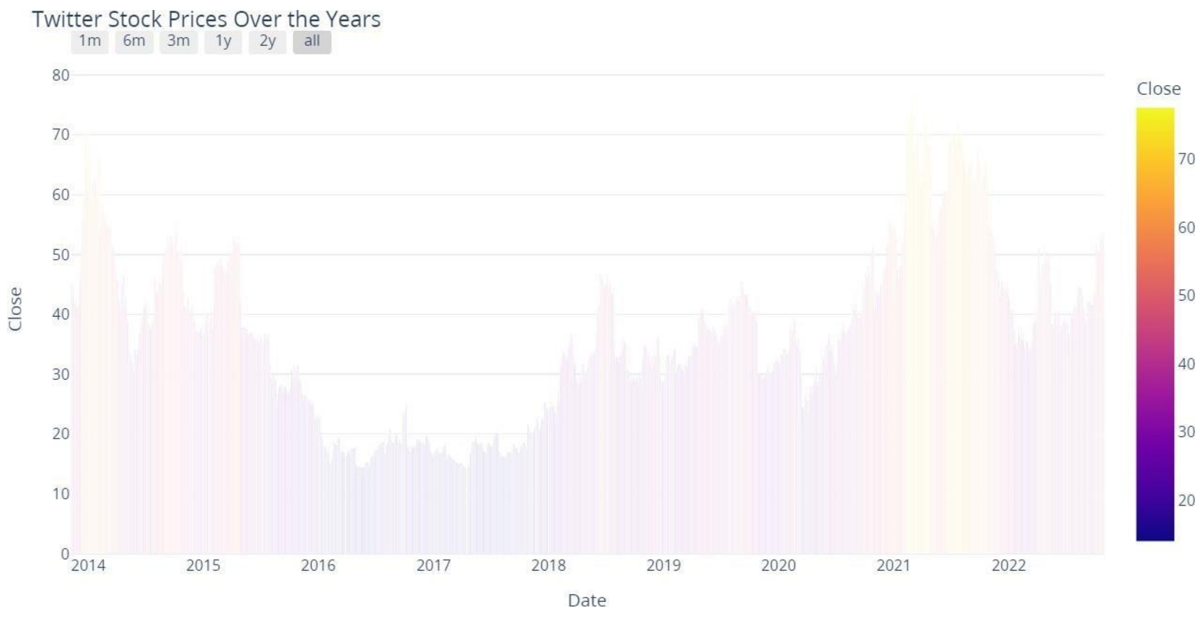
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Twitter Stock Prices Over the Years

Complete Timeline of Twitter




Twitter Stock Prices Over the Years
1m 6m 3m 1y 2y all
Close
Date
Close
0
10
20
30
40
50
60
70
80
2014
2015
2016
2017
2018
2019
2020
2021
2022
20
30
40
50
60
70


Twitter Stock Prices Over the Years
1m 6m 3m 1y 2y all
Close
Date
Close
0
10
20
30
40
50
60
70
80
2014
2015
2016
2017
2018
2019
2020
2021
2022
20
30
40
50
60
70

7.Conclusion

Overall, the examination led to utilizing Python on Twitter financial exchange information uncovers a few significant experiences.

First and foremost, the stock's unpredictability has all the earmarks of being impacted by tweets from high-profile people as well as significant occasions and news. These tweets' sentiment analysis enables a deeper comprehension of market sentiment and its influence on stock performance.

Besides, the connection investigation shows areas of strength for a connection between the quantity of tweets and the stock's cost development. This proposes that Twitter movement can be a dependable indicator of the financial exchange's way of behaving.

Besides, the utilization of AI calculations, for example, LSTM models, can successfully figure future stock costs in view of authentic tweet information. This connotes the possible benefit of using web-based entertainment information in foreseeing financial exchange patterns.

Finally, itis crucial for note that while Twitter information gives significant bits of knowledge, it ought to be utilized related to other essential and specialized examination methods for a complete assessment of financial exchange execution.

Overall, the research backs up the idea that Twitter can be a useful tool for analyzing and forecasting the stock market because it provides unique and immediate insights in to investor sentiment and market trends. Further examination and investigation of cutting edge methods can give more exact expectations and upgrade venture dynamic cycles.

References

1. Chen, C., Zhang, D., Li, X., & Pan, Y. (2020). Sentiment analysis of Twitter data for predicting stock market movements. Journal of BigData,7(1), 1-16.

2. Alqahtani, M., & Adedeji, A. (2021). Stock market prediction using sentiment analysis of Twitter data. Journal of Information Science, 47(2), 232-253.

3. Abualigah, L. M., Khedher, W. B., & Jayaraman, R. (2020). Stock market prediction using natural language processing on Twitter data. Journal of Financial Innovation, 6(1),1-19.

4. Belem, F.,&deOliveira,V.M.(2015).SentimentanalysisonTwitterforstockmarketprediction.ExpertSystems withApplications,42(24),9603-9611.

5. Zhang, T., Zhang, Y., Xia, H., & Chen, T. (2020). Stock market volatility forecasting usingsentimentanalysisofTwitterdata.NeuralProcessingLetters,51(1), 621-640.

6. Rilett, L. R., Kanhere, N. S., & Desai, K. P. (2019). Twitter sentiment analysis for stock market prediction using word2vec and random forest classifier. Journal of Artificial Intelligence and DataAnalysis,3(1),11-20.

7. Ebrahimi, N., & Shavvalpour, S. (2020). Stock market prediction using Twitter sentiment analysis and machine learning techniques. Journal of Data Science and Applications,3(2),37-48.

8. Ahmed,L.,Majumder,N.,&Mustafa,N.(2021).TwittersentimentanalysisforstockmarketpredictionusingLSTMandBERT.JournalofBigDataAnalyticsinTransportation,3(1), 1-17.

9. Li,X.,Chen,C.,Zhang,D.,&Pan,Y.(2021).StockmarketpredictionusingsentimentanalysisofTwitterdata based on machine learning methods. Sustainability, 13(7),3931.

10. Singh,H.,Sharma,S.,&Jain,S.(2020).Stockmarketpredictionusingsentimentanalysisandmachinelearning. Procedia ComputerScience, 171,420-426