STOCK ANALYSIS AND FORECASTING BASED ON TWEETS

Divya Sai Gummadi
Computer Science
University of Missouri-Kansas City
Missouri State, Kansas City
dgpbb@umsystem.edu

Sai Prasanna Vetapalem
Computer Science
University of Missouri-Kansas City
Missouri State, Kansas City
pvkg8@umsystem.edu

Abhinay Reddy polimera
Computer Science
University of Missouri-Kansas City
Missouri State, Kansas City
apxgx@umsystem.edu

Sivendra Pammi
Computer Science
University of Missouri-Kansas City
Missouri State, Kansas City
spfk7@umsystem.edu

Hari Kiran Reddy Muthyala
Computer Science
University of Missouri-Kansas City
Missouri State, Kansas City
hm27w@umsystem.edu

Madhav Kilaru
Computer Science
University of Missouri-Kansas City
Missouri State, Kansas City
mkbn3@umsystem.edu

ABSTRACT

Financial engineering considers stock market forecasting to be a key topic, and studies on the subject are becoming more and more valuable. Stock market is always tough to analyse with graphs, news, important dates, and social media. These creates inaccuracies in investing the stocks. In this project, we investigate the stock market and twitter data are used to examine the fluctuations using the stacked LSTM model. It collects information from Twitter about stock symbols to analyse and forecast stock movement. In this study, we gather and process a twitter data set to analyse the stock market system and the effect of public opinion on firm market value. The data was then subjected to sentiment analysis, and the trend for the tweets with the highest engagement was then predicted. Then, to help with stock movement forecasts, we deployed a stacked LSTM model.

1. INTRODUCTION

People investments in stock market are increasing rapidly, but many of them fail to analyse the stock fluctuations, so it is important issue to address. The biggest issue facing researchers today involves forecasting the stock market price due to the complexity involved in determining an exact value that can match the real stock price.

Stock market prediction is the technique of estimating the future value of a specific business's stock to provide investors with an estimate of gain or loss when investing in that specific company stock. Twitter is a key social media platform for stock price forecasting. On the social media site Twitter, millions of tweets are sent every day. The stock market is covered in newspaper headlines as well, and this information can be used to make predictions. One can perform a prediction procedure using the data from Twitter.

Historical data have been utilised extensively in prior research on stock prediction. When conducting research with historical data, technical analysis is one method that can be used to predict future stock market patterns and prices. On historical stock price data, researchers employed a variety of machine learning approaches, including regression analysis and deep learning. These research, however, omitted outside variables like social media. Since it is believed that prices change because of human behaviour, which can be represented in social media, it is crucial to make use of social media data since events conveyed through social media can greatly affect stock prices and trends.

Big data tools make it easier to analyse vast amounts of data, producing more effective results. This project mines data about a symbol from Twitter stocks, then analyses and predicts the movement of the stock using that data.

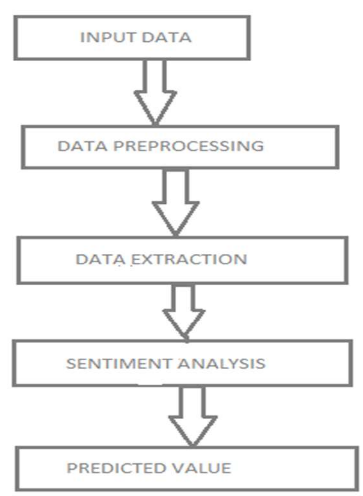


Figure 1. High level structure.

KEYWORDS

Sentiment Analysis, Vader sentiment, Stock Prediction, LSTM, NumPy, PySpark, Pandas, Matplotlib.

2. RELATED WORK

According to the microeconomics concept, the main reason why stock prices change is due to supply and demand in the market. Because of the non-stationarity, non-linearity, and noise in the environment, which in turn affects the unpredictability of stocks, forecasting stock values is difficult. The predicted value of a stock depends on a wide range of variables, including market conditions, government stability, customer value, client feedback, trader expectations, and social media.

The Elliot wave hypothesis, which claims that some stock movements repeat over time, is one theory that can be used to analyse market trends aside from mathematical patterns. But different theories exist. Malkiel believes that stocks can be explained as having a random walk, meaning that due of their irrational occurrence, future prices cannot be anticipated by prior data. Recent studies have demonstrated that stock movement is predictable, and this thesis consequently assumes that it is.

Agarwal and Apoorv investigate different machine learning methods for evaluating a tweet's positivity or negativity. The author employs a variety of methods, including support vector machines and naive bayes. Support vector machine techniques like the Naive Bayes classifier, which was used to assess sentiment in the tweet data set, would be utilised to forecast market movement.

The study highlighted by A. E. O. Carosia, G. P. Coelho, and A. E. A. Silva, SM actions that, because of domestic and international factors, have a significant impact on the market value of specific enterprises. Three aspects of Brazilian social media behaviour on Twitter were examined in the study: It is important to recognise the following: overall number of Tweet emotions; Tweet sentiments with likes; and Balanced sentiments.

News SA was presented by the author P. G. S. Mate. A business that tracks stock market volatility was the one that predicted the stock price. Additionally, they suggested that SA be used to rate articles by joining together ratings that were good, negative, and neutral into a single concatenated string. Any classification model for stock market predictions examines the SA's output.

D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, created a TCS market price prediction model based on the start, major, minor, finish, and quantity variables. The study investigated the impact of regression models based on optimistic estimations of either the anticipated outcomes on linear, polynomial, and radial base functions. The linear regression strategy outperformed other methods with a confidence level of 0.97. As the global market economy expands, stock values fluctuate significantly. Even with experience, it is challenging to estimate stock prices due to historical trends and stock valuations. Since the beginning, stock predictions have been made using a variety of techniques, either out of pure avarice or pure curiosity about the future. Some use ML models, some use sentiment analysis, and some use a combination of the two. Previous studies have discovered a connection between market fluctuations and social media elements like news, trending topics online, and the tone of posts. By incorporating input from social media and historical stock data into the LSTM Model, forecasts can be expected more precisely than when using solely past market data.

3. PROCESSING STEPS

This methodology incorporates a few standard procedures. The following are those:

3.1 DATA COLLECTION

On Kaggle, historical stock price data is available. Price information for the selected stock markets is downloaded in csv file format from Kaggle for the selected time frame. The seven features of the downloaded data files—Date, Open, High, Low, Close, Volume, and Adjusted Close—display, for a given date, the stock traded day, stock open price, stock maximum trading price, stock lowest trading price, stock closing price, number of shares traded, and stock closing price when dividends are paid to investors, in that order. Only the Date and Close price are used in this report.

We have collected four datasets from Kaggle related to company, company ticker symbols, tweets, stock prices. We have collected tweet data of top companies: Apple, Tesla, Google, Amazon Over 1.6 million tweets are collected. For over 10-year period from 2011 to 2021.

tweet_id	writer	post_date	body	comment_num	retweet_num	like_num
550441509175443456	VisualStockRSRC	1420070457	lx21 made $10,008...	0	0	1
550441672312512512	KeralaGuy77	1420070496	Insanity of today...	0	0	0
550441732014223360	DozenStocks	1420070510	S&P100 #Stocks Pe...	0	0	0
550442977802207232	ShowDreamCar	1420070807	$GM $TSLA: Volksw...	0	0	1
550443807834402816	i_Know_First	1420071005	Swing Trading: Up...	0	0	1
550443808606126081	aaplstocknews	1420071005	Swing Trading: Up...	0	0	1
550443809700851716	iknowfirst	1420071005	Swing Trading: Up...	0	0	1
550443857142611968	Cprediction	1420071016	Swing Trading: Up...	0	0	1
550443857595600896	iknowfirst_br	1420071017	Swing Trading: Up...	0	0	1
550443857692078081	Gold_prediction	1420071017	Swing Trading: Up...	0	0	1
550443858010861568	IKFResearch	1420071017	Swing Trading: Up...	0	0	1
550444112328261632	GetAOM	1420071077	$UNP $ORCL $QCOM ...	0	0	0
550444969924653056	AppleNewsAAPL	1420071282	$AAPL Apple goes ...	0	0	1
550444970738335744	espositooooo	1420071282	"@WSJ: Apple is b...	0	0	0
550445066444369921	Bidnessetc	1420071305	Apple filed for i...	0	0	0
550445850170642432	JorelLaraKalel	1420071492	@CNBC 15 Top #tra...	0	0	2
550447574285418497	btcgemini	1420071903	We searched throu...	0	0	0
550447850857828352	JorelLaraKalel	1420071969	Top 10 searched #...	0	0	2
550447998577426433	UPBOptionMil	1420072004	2014 The Year in ...	0	2	2
550448085789200384	MacHashNews	1420072025	Give your brain a...	0	0	0

Figure 2: Tweet data of all companies

Our dataset consists of 7 columns namely

1.Tweet id 2. Writer 3. Post_Date 4. Body 5. Comment_num 6. Retweet_num 7. Like_num

tweet_id	ticker_symbol
550803612197457920	AAPL
550803610825928706	AAPL
550803225113157632	AAPL
550802957370159104	AAPL
550802855129382912	AAPL
550802745737768960	AAPL
550797494188142592	AAPL
550797275786518528	AAPL
550797272686923776	AAPL
550796617444765696	AAPL
550795512899960832	AAPL
550795254102638593	AAPL
550795167318700033	AAPL
550795088821886976	AAPL
550793298357391360	AAPL
550793247669231616	AAPL
550793108242198528	AAPL
550791919815892992	AAPL
550791232738590720	AAPL
550790423888035840	AAPL

Figure 3: Tweet ID and its associated ticker symbols

ticker_symbol	company_name
AAPL	apple
GOOG	Google Inc
GOOGL	Google Inc
AMZN	Amazon.com
TSLA	Tesla Inc
MSFT	Microsoft

Figure 4: Company and its ticker symbols

3.2 PRE-PROCESSING

The following rules will be used to preprocess each tweet. By applying the following rules to all the text, the preparation of the data will be carried out. For the sake of our study, we converted the post date column from the dataset into date format.

tweet_id	writer	body	comment_num	retweet_num	like_num	ticker_symbol	total_engagements	t_date
692169663577485315	ValaAfshar	Apple has $216 bi...	42	984	677	AAPL	1703.0	2016-01-27
770310550991605760	cnntech	Apple's next iPho...	11	729	918	AAPL	1658.0	2016-08-29
575014851363405824	RANsquawk	Loving my Apple W...	66	882	654	AAPL	1602.0	2015-03-09
816359802733555712	DavidSchawel	Sometimes hard to...	14	646	900	AMZN	1560.0	2017-01-03
854690001866686464	philstockworld	Will We Hold It W...	0	969	520	AMZN	1489.0	2017-04-19
854690001866686464	philstockworld	Will We Hold It W...	0	969	520	TSLA	1489.0	2017-04-19
1021481848403382272	QTRResearch	Guys - I'm beside...	207	317	899	TSLA	1423.0	2018-07-23
875518367003791362	SJosephBurns	$AMZN has no stor...	40	509	837	AMZN	1386.0	2017-06-16
862303523010203648	philstockworld	Watergate Wednesd...	1	971	400	TSLA	1372.0	2017-05-10
613718497219076096	Carl_C_Icahn	Sold last of our ...	153	671	533	AAPL	1357.0	2015-06-24
1054728662786826240	CitronResearch	$TSLA dropping ea...	148	308	861	TSLA	1317.0	2018-10-23
1018938697415315457	epichedge	Live view of $AMZ...	7	366	927	AMZN	1300.0	2018-07-16
1020077355346169857	vincent13031925	"Tesla Spokespers...	38	256	986	TSLA	1280.0	2018-07-19
1199424478536753155	AlexSibila	~Tesla feature re...	563	48	662	TSLA	1273.0	2019-11-26
1135604016015060993	willchamberlain	FACEBOOK, GOOGLE ...	58	389	826	GOOG	1273.0	2019-06-03
1045404879341137921	Reuters	SEC files lawsuit...	56	630	585	TSLA	1271.0	2018-09-27
1167316598283071495	TeslaNY	.@Tesla Model 3 p...	36	279	952	TSLA	1267.0	2019-08-30
1020036769629143040	Microsoft	$MSFT Q4 EARNINGS...	26	333	896	MSFT	1255.0	2018-07-19
1209424426904801280	YCalenge	Last night, $TSLA...	234	172	835	TSLA	1241.0	2019-12-24
1050135192109760525	charliebilello	% Below 52-week h...	34	415	783	AMZN	1232.0	2018-10-10

Figure 5: adding total engagement columns to filter most engaged tweets.

We included a new column called Total Engagement during data preparation, which is defined as the sum of tweet comments, likes, and retweets. To filter out tweets that did not attract a lot of attention and are therefore likely to have a smaller impact, we anticipated that the overall engagement value would be more than 200.

3.3 SENTIMENT ANALYSIS

Sentiment analysis is the process of extracting conceptual meaning and deciphering implicit information, such as subjective information, in a text. The retrieved data can then be included into statistical or machine learning models. The text itself has been the main subject of sentiment research on social media.

Algorithms execute sentiment analysis using a systematic manner to extract things like polarity, subjects, and opinions from the text. Two approaches that can be used are rule-based language modelling and artificial intelligence-based computation of hidden patterns. Since the grammar and structure of language are difficult to summarize and express with computing models, sentiment analysis is seen as being complex. One of the issues is sarcasm detection and word ambiguity. Depending on the situation, words can signify multiple things. Additionally, social media communication is condensed and uses emoticons, shortenings, and uppercase characters to express intent and feelings. demonstrated how social media texts might represent emotions differently from normal language, which makes a harmonized sentiment analysis more difficult.

We categorized tweets into three types: Positive, Negative and Neutral. We then used Vader sentiment. Valence Aware Dictionary for Sentiment Reasoning is a text sentiment analysis model that is sensitive to both emotion polarities positive and negative and intensity of emotion to determine sentiment of each post.

Neutral tweets were removed because they had no impact on the stock price. There are 10448 negative sentiments and 17586 positive sentiments, as seen

we shall now make a function that takes in text and returns the sentiment of that text.

```
[16] def getSentiment(body):

        from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
        analyzer = SentimentIntensityAnalyzer()

        assert body is not None
        vs = analyzer.polarity_scores(body)
        score = vs['compound']

        if (score >= 0.05):
            return "Positive"

        elif (score < 0.05 and score > -0.05):
            return "Neutral"

        elif (score <= -0.05):
            return "Negative"

        print(score)
```

Figure 6: Sentiment analysis of tweets

tweet_id	writer	body	comment_num	retweet_num	like_num	ticker_symbol	total_engagements	t_date	Score
3577485315	ValaAfshar	Apple has $216 bi...	42	984	677	AAPL	1703.0	2016-01-27	Neutral
0991605760	cnntech	Apple's next iPho...	11	729	918	AAPL	1658.0	2016-08-29	Neutral
1363405824	RANsquawk	Loving my Apple W...	66	882	654	AAPL	1602.0	2015-03-09	Positive
2733555712	DavidSchawel	Sometimes hard to...	14	646	900	AMZN	1560.0	2017-01-03	Negative
1866686464	philstockworld	Will We Hold It W...	0	969	520	AMZN	1489.0	2017-04-19	Neutral
1866686464	philstockworld	Will We Hold It W...	0	969	520	TSLA	1489.0	2017-04-19	Neutral
8403382272	QTRResearch	Guys - I'm beside...	207	317	899	TSLA	1423.0	2018-07-23	Negative
7003791362	SJosephBurns	$AMZN has no stor...	40	509	837	AMZN	1386.0	2017-06-16	Negative
3010203648	philstockworld	Watergate Wednesd...	1	971	400	TSLA	1372.0	2017-05-10	Neutral
7219076096	Carl_C_Icahn	Sold last of our ...	153	671	533	AAPL	1357.0	2015-06-24	Positive
2786826240	CitronResearch	$TSLA dropping ea...	148	308	861	TSLA	1317.0	2018-10-23	Negative
7415315457	epichedge	Live view of $AMZ...	7	366	927	AMZN	1300.0	2018-07-16	Neutral
5346169857	vincent13031925	"Tesla Spokespers...	38	256	986	TSLA	1280.0	2018-07-19	Negative
6015060993	willchamberlain	FACEBOOK, GOOGLE ...	58	389	826	GOOG	1273.0	2019-06-03	Negative
8536753155	AlexSibila	~Tesla feature re...	563	48	662	TSLA	1273.0	2019-11-26	Positive
9341137921	Reuters	SEC files lawsuit...	56	630	585	TSLA	1271.0	2018-09-27	Negative
8283071495	TeslaNY	.@Tesla Model 3 p...	36	279	952	TSLA	1267.0	2019-08-30	Neutral
9629143040	Microsoft	$MSFT Q4 EARNINGS...	26	333	896	MSFT	1255.0	2018-07-19	Negative
6904801280	YCalenge	Last night, $TSLA...	234	172	835	TSLA	1241.0	2019-12-24	Positive
2109760525	charliebilello	% Below 52-week h...	34	415	783	AAPL	1232.0	2018-10-10	Neutral

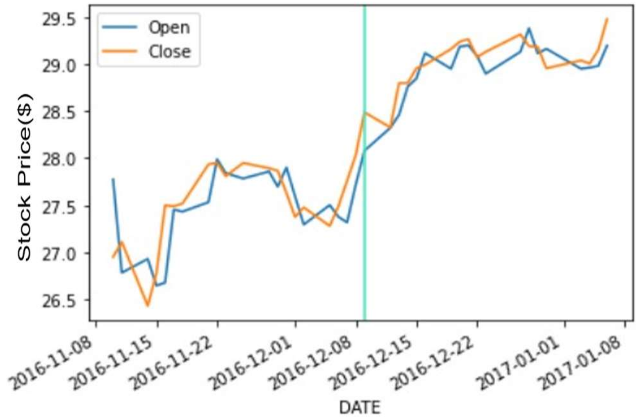
Figure 7: Tweets with positive, negative, and neutral Sentiments

4. RESULTS

We projected the results from the Real-Time Stock Market data set of the four companies that we collected earlier. After performing sentiment analysis, we filtered out Neutral tweets as they do not have any influence on the stock market. We analyzed the stock prices of various companies with the help of sentiment analysis performed on the tweets. We have defined a python function for plotting the different values and then to compare the stock price based on the sentiment obtained on that day based on the tweets and when passed the date, ticket symbol we can visualize the values as shown in the next pages.

ticker_symbol	Date	Open	High	Low	Close	Adj Close	Volume
AAPL	11/4/2022	142.089996	142.669998	134.380005	138.380005	138.380005	140716700
AAPL	11/3/2022	142.059998	142.800003	138.75	138.880005	138.650009	97918500
AAPL	11/2/2022	148.949997	152.169998	145	145.029999	144.78981	93604600
AAPL	11/1/2022	155.080002	155.449997	149.130005	150.649994	150.400497	80379300
AAPL	10/31/2022	153.160004	154.240005	151.919998	153.339996	153.086044	97943200
AAPL	10/28/2022	148.199997	157.5	147.820007	155.740005	155.482086	164762400
AAPL	10/27/2022	148.070007	149.050003	144.130005	144.800003	144.560196	109180200
AAPL	10/26/2022	150.960007	151.990005	148.039993	149.350006	149.102661	88194300
AAPL	10/25/2022	150.089996	152.490005	149.360001	152.339996	152.087708	74732300
AAPL	10/24/2022	147.190002	150.229996	146	149.449997	149.202484	75981900
AAPL	10/21/2022	142.869995	147.850006	142.649994	147.270004	147.026108	86464700
AAPL	10/20/2022	143.020004	145.889999	142.649994	143.389999	143.152527	64522000
AAPL	10/19/2022	141.690002	144.949997	141.5	143.860001	143.62175	61758300
AAPL	10/18/2022	145.490005	146.699997	140.610001	143.75	143.511932	99136600
AAPL	10/17/2022	141.070007	142.899994	140.270004	142.410004	142.174164	85250900
AAPL	10/14/2022	144.309998	144.520004	138.190002	138.380005	138.150833	88512300
AAPL	10/13/2022	134.990005	143.589996	134.369995	142.990005	142.753204	113224000
AAPL	10/12/2022	139.130005	140.360001	138.160004	138.339996	138.110886	70433700
AAPL	10/11/2022	139.899994	141.350006	138.220001	138.979996	138.749832	77033700
AAPL	10/10/2022	140.419998	141.889999	138.570007	140.419998	140.187439	74899000

Figure 8: Real-Time Stock Market Data

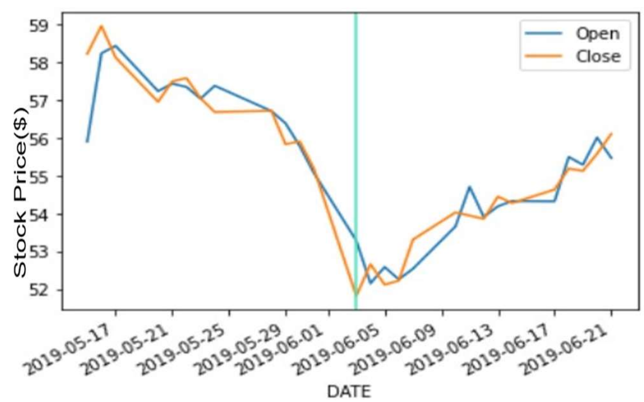



```
[ ] df3_f.filter(df3_f.t_date == '2016-12-09').filter(df3_f.ticker_symbol == 'AAPL').show()
```

tweet_id	writer	body	comment_num	retweet_num	like_num	ticker_symbol	total_engagements	t_date	Score
807233825659097088	philstockworld	Thank Trump It's ...	1	911	33	AAPL	945.0	2016-12-09	Positive

Figure 9: Apple Stock Trend

Looking at the price of **AAPL** on 12-09-2016, it appears that the price has a **positive rise** ant the outcome of sentiment analysis on that day is positive. So here we can conclude there is an effect of stock price based on the tweets.

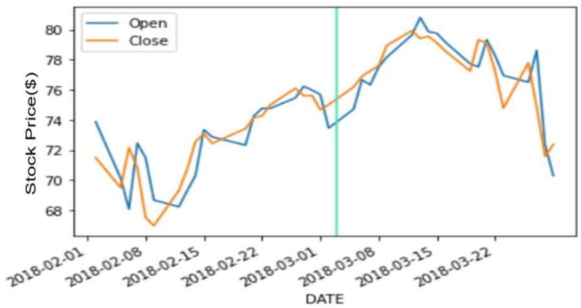



```
[ ] df3_f.filter(df3_f.t_date == '2019-06-03').filter(df3_f.ticker_symbol == 'GOOG').show()
```

tweet_id	writer	body	comment_num	retweet_num	like_num	ticker_symbol	total_engagements	t_date	Score
1135604016015060993	willchamberlain	FACEBOOK, GOOGLE ...	58	389	826	GOOG	1273.0	2019-06-03	Negative
1135658420785336320	matthewstoller	The House Judicia...	8	38	93	GOOG	139.0	2019-06-03	Negative
1135638699754938368	GretaLWall	#MarketWrap Major...	2	16	47	GOOG	65.0	2019-06-03	Negative

Figure 10: Google Stock Trend

In case of Google, **negative sentiment** appears to have taken effect on the date 06-03-2019 and the price appears to have fallen, somewhat deviating from the pattern.




```
[ ] df3_f.filter(df3_f.t_date == '2018-03-03').filter(df3_f.ticker_symbol == 'AMZN').show()

+------------------+---------------+--------------------+-----------+-----------+--------+-------------+-----------------+----------+--------+
|          tweet_id|         writer|                body|comment_num|retweet_num|like_num|ticker_symbol|total_engagements|    t_date|   Score|
+------------------+---------------+--------------------+-----------+-----------+--------+-------------+-----------------+----------+--------+
|969982565787545601|victoriouscake|Mutual funds are ...|          2|          8|      61|         AMZN|             71.0|2018-03-03|Positive|
+------------------+---------------+--------------------+-----------+-----------+--------+-------------+-----------------+----------+--------+
```

Figure 11: Amazon Stock Trend

Looking at the price of AMZN on 03-03-2018, it appears that there has been immediate rise in the price after the most engaged tweets were made**.**

5. STOCK PREDICTION & FORECASTING

In predicting the future trend using time series analysis, the scale of the data is very important. Here the close value is ina kind of scale, so we should seek to transform the value. The values will be transformed from 0 to 1 using the min- max scalar. We should reshape so that fit transform can be used.

LSTM:

“Long Short - Term Memory (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory”.

There are three gates in the LSTM: The input gate is a devicethat contributes information to the state of a cell. The forget gate eliminates information from the model that is no longerneeded. The LSTM output gate selects the information to be displayed as output unpredictable circumstances, price movement may not necessarily follow the historical trend. To support our investing decision-making, extra fundamental and market analysis is necessary in this case.

The major drawback of adopting any machine learning system to forecast stock prices is that we can only run a back test on previous data, and under various

We used the Apple Stock Data for predicting the future trends using LSTM model.

	Date	Low	Open	Volume	High	Close	Adjusted Close
0	12-12-1980	0.128348	0.128348	469033600.0	0.128906	0.128348	0.100922
1	15-12-1980	0.121652	0.122210	175884800.0	0.122210	0.121652	0.095657
2	16-12-1980	0.112723	0.113281	105728000.0	0.113281	0.112723	0.088636
3	17-12-1980	0.115513	0.115513	86441600.0	0.116071	0.115513	0.090830
4	18-12-1980	0.118862	0.118862	73449600.0	0.119420	0.118862	0.093463
...	...	...	...	...	...	...	...
10167	12-04-2021	130.630005	132.520004	91420000.0	132.850006	131.240005	131.240005
10168	13-04-2021	131.929993	132.440002	91266500.0	134.660004	134.429993	134.429993
10169	14-04-2021	131.660004	134.940002	87222800.0	135.000000	132.029999	132.029999
10170	15-04-2021	133.639999	133.820007	89347100.0	135.000000	134.500000	134.500000
10171	16-04-2021	133.279999	134.300003	84818500.0	134.669998	134.160004	134.160004

10172 rows × 7 columns

Figure 12: Apple Stock Data

```
[ ] model.summary()

Model: "sequential_5"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_15 (LSTM)               (None, 100, 50)           10400
_________________________________________________________________
lstm_16 (LSTM)               (None, 100, 50)           20200
_________________________________________________________________
lstm_17 (LSTM)               (None, 50)                20200
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 51
=================================================================
Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0
_________________________________________________________________
```

Figure 13: LSTM model summary

We should always restructure our X train in 3-D and add 1 when implementing any LSTM. The reason for this is the time step and the 1 is supplied to the LSTM. Here we are using a sequential model and adding the layers of the LTSM.

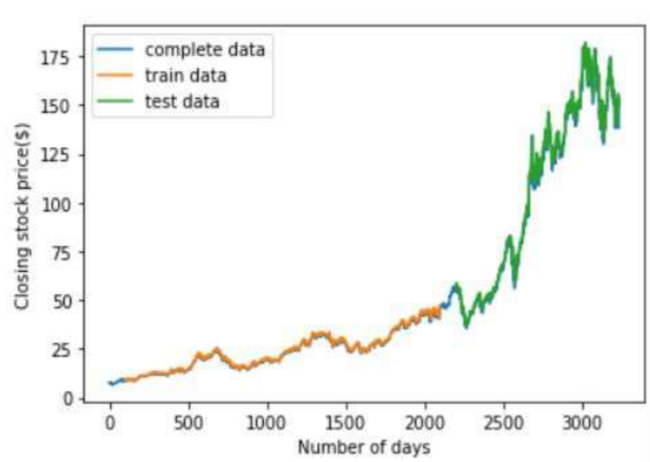


Figure 14: Apple Trend

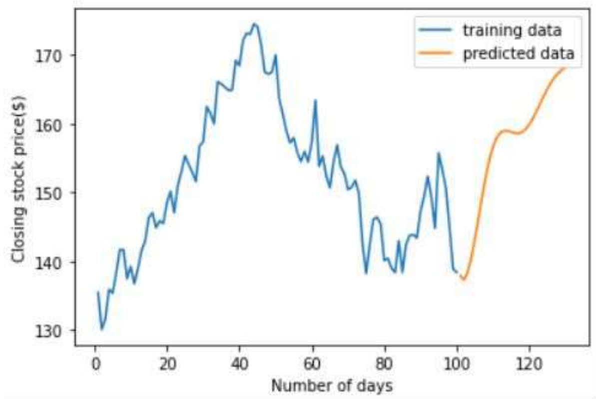


Figure 15: Predicting the Trend for next 30 days

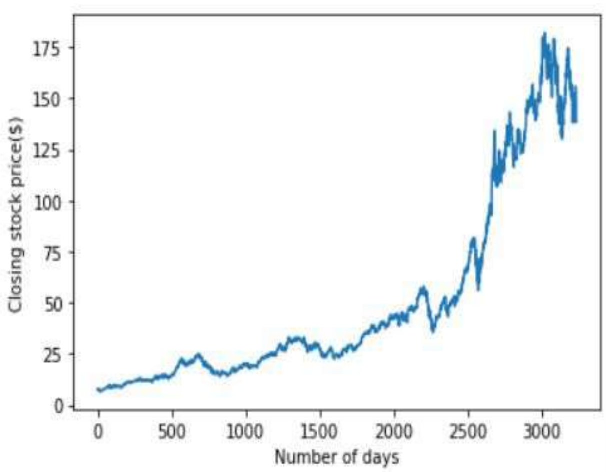


Figure 16: Original Graph

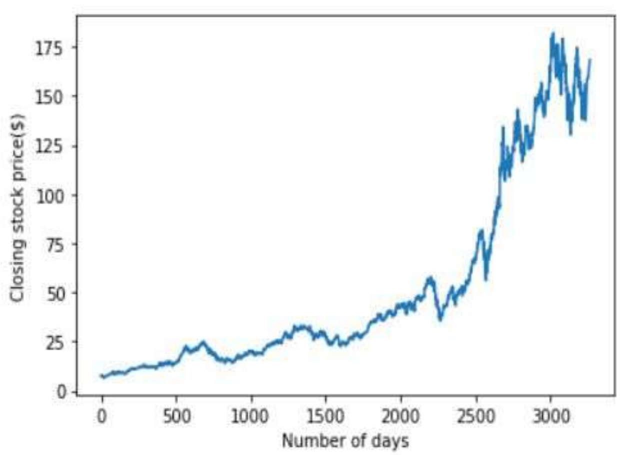


Figure 17: Extended Graph

Here, we have predicted the stock price for next 30 days based on the training set, compared, and extended it to original graph using matplotlib libraries.

6. CONCLUSION AND FUTURE WORK

In this study, we investigated the relationship between sentiment analysis of Twitter data and stock market price forecasts for all the companies that were included. The outcome of the prediction procedure makes it abundantly evident that we have obtained a correct value that appropriately corresponds to the current stock price.

Out of 8 random tweets only 4 (50%) influenced the stock prices when compared with the opening and closing values. So, in conclusion, yes, twitter does influence the stock market, but it is not necessary that if your tweet does go viral, based on emotion behind tweet it will influence the stock prices of company.

While the exact price points from our predicted price weren't always close to the actual price, our model did still indicate overall trends such as going up or down. This project teaches us how the LSTMs can be effective in times series forecasting.

Future work on this study will involve applying the models to various international stock exchanges. Additionally, using data that spans more than a year may yield conclusions that are more accurate. Additionally, studying the models in various economic climates, such as booms or recessions, may help us better understand the models' productivity. Additionally, using a neural network to classify tweets based on sentiment analysis may produce superior results.

7. AUTHOR CONTRIBUTIONS

Abhinay: Using widely shared sentiments from Sentiment Analysis, he projected the trend for all four companies.

Hari Kiran: Contributed to the data set collection and performed sentiment analysis on all the tweets with the highest engagement.

Sai Prasanna: Applied LSTM to predict the stock trend for the upcoming 30 days and to analyze test and train data.

Divya: Data preprocessing including removing unwanted values, w.r.t data, and duplicate tweets and casting columns as required.

Sivendra: The gathering and analysis of data from many sources.

Madhav: Supported sentiment analysis and trend projection for popular sentiment.

8. REFERENCES

1. Ching-Ru Ko and Hsien-Tsung Chang LSTM-based sentiment analysis for stock price forecast https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7959635/

2. a. Mittal and a. Goel. "Stock Prediction Using Twitter Sentiment Analysis." Tomx.Inf. Elte.Hu, (June), 2012.

3. Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." Proceedings of the Workshop on Languages in social media. Association for Computational Linguistics, 2011.

4. a. e. o. crosier, g. p. Coelho, and a. e. a. silva, "analyzing the Brazilian financial market through Portuguese sentiment analysis in social media," appl. artif. intel., vol. 34.

5. Kaz Anova, M.M. (2017) Sentiment140 Dataset with 1.6 million Tweets. https://www.kaggle.com/kazanova/sentiment140

6. Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." Proceedings of the Workshop on Languages in social media. Association for Computational Linguistics, 2011.

7. p. g. s. mate, "issn no: 1006-7930 stock prediction through news sentiment analysis.," vol. xi, no. viii, pp. 36–40, 2019.

8. d. Bhuiya, g. Kaushal, a. Sharma, and u. Singh, "stock market predication using a linear regression," proc. int. conf. electron. common. aerosp. technol. icecap 2017, vol. 2017-janua.