

# PREDICTING STOCK PRICES USING TWEET FREQUENCY AND AI: LEVERAGING SOCIAL MEDIA INSIGHTS TO FORECAST TOMORROW'S MARKET TRENDS

Zimo Liu<sup>1</sup>, Lin Yang<sup>2</sup> and Tami Takada<sup>3</sup>

<sup>1</sup>Beijing No. 12 High School, No. 15 Yize Road Fengtai District, Beijing

<sup>2</sup>13711 Somerset Ln SE, Bellevue WA 98006

<sup>3</sup>Computer Science Department, California State Polytechnic University, Pomona, CA 91768

## **ABSTRACT**

*I use the frequency of the tweet counts of the stock ticker to predict the stock price [1]. Using AI to predict the price with today's tweet count and the highest and lowest stock price to predict tomorrow's stock price [2]. The benefits of predicting stock prices are minimizing losses and getting a better idea about making money. But since the price of stocks is hard to predict, analyzing the number of price tickers discussed in social media can help people to know more precisely trends. "They analyzed the activity of 150 companies chosen at random from the S&P 500 Index and noticed a correlation between the number of tweets they sent and their share price [3]. Having observed this, the researchers devised a mathematical model, applied it to an imaginary portfolio and outperformed other financial strategies based on financial analysis by as much as 11%" [4]. This indicates the reliability of predicting stock prices with social media. Compared with some popular methods such as Auto-Regressive Conditional Heteroscedastic or Generalized Auto-Regressive Moving Average, using social media to predict price can consider the effect of social trends on the stock price [5]. Using social media to predict the stock price is not subjective and is much cheaper to computerize.*

## **KEYWORDS**

*Stock Price Prediction, Tweet Frequency Analysis, AI-Based Forecasting, Social Media Trends*

## **1. INTRODUCTION**

Using social media to predict stock prices can help people earn more money. Because the stock market is risky, people need more methods to predict the price [6]. If people get the relationship between social media and stock price, they can make better decisions enabling them to make more money and lose less. "A new study conducted by a doctoral student at Pace University, in association with Famecount (which tracks how popular brands are according to social media) concludes that social media popularity can reliably predict daily stock prices." The prices of stocks are not logically moving and there are already some methods and solutions for people to predict the prices. But an external factor that people can use to predict the stock price is beneficial and especially when social media is the place where everyone can share their ideas about different stock tickers. Therefore, my topic can help people to link social media with the stock market and find the relationships between social media and stock prices. "If traders can adequately predict the stock trends and patterns, they can earn a considerable profit margin.

However, stock markets are very volatile and, consequently, difficult to predict. External factors, such as social media and financial news, can have widespread effects on stock price movement.”

Some prediction techniques and systems have been proposed to predict stock prices, which allows the user to make a better decision in the risky stock market. However, these proposals assume that every person can understand the methods easily and can practice in the real market. Their implementations are also limited in scale, with samples given for some traditional econometric models, such as the regression models and ARIMA [7]. Other techniques, such as the MACD [8]. It is only useful for short term traders instead of long run traders because it only considers the short run average moving average which only can predict the short run trend of the stock. Also, the method and algorithm used cannot be too sophisticated because often results are hard to practice. A second practical problem is that some users find the method is really hard to understand, including using machine learning. Because the stock price can not only be predicted by a single method, if people want to predict a complex stock trend, they need to use many methods. Thus, people need a method which is easier to practice and can provide enough prediction and information about the stock price. Compared with using social media to predict stock prices, some current famous indicators, such as OBV, RSI or average direction index are always focused on the market itself rather than link the stock market to other factors in the real world which can influence the stock prices [9].

Our goal is to predict the stock price with the tweet counts of the stocks. Our method is inspired by the characteristic of the stock market which is risky, and people are continually interested in predicting price. Some researchers also have practiced that using social media to predict the stock price is useful and reliable, therefore I got the idea of predicting the stock price with social media. There are some good features of predicting stock price with social media. First, social media is the place where everyone can talk and share their ideas with others. People are easily influenced by others' opinions, especially when this decision is about making money. Therefore, the link between social media activity and the stock price is really worthwhile to find. Second, different stocks have different stock trends. Some stocks will raise price tomorrow if the tweet counts increase and some stocks will decrease price tomorrow if the tweet counts increase. Therefore, we need to use AI to help us to predict accurate results of stock prices by considering stock's unique trend. Third, my project can predict the price of the stock, which is easy to understand and can directly use the number provided to make a decision in the stock market. Therefore, we believe that our method is easier to understand and practical.

The experiments were conducted over three days, with a focus on collecting and comparing predicted stock prices for three selected stocks. Five data points were used for comparison, including the highest price, lowest price, open price, closed price, and whether the predicted price fell within the range of actual prices. The experiments aimed to assess the reliability of the prediction app, with each experiment repeated three times to ensure sufficient sample size.

Based on the results of all experiments, it was found that the app provided reliable data for consideration. The maximum difference between the predicted price and the real price was less than 3, indicating a relatively accurate prediction. Moreover, the majority of stocks in the experiments fell within the range of real stock prices, suggesting the app's ability to capture the overall trend of the market. It is important to note that stock market prices are subject to constant fluctuations during trading hours, and the comparison focused on the closing price, which served as the endpoint indicator for the experiment.

While the predicted prices were not an exact match for the real closed prices, the small difference of less than 3 demonstrates the reliability of the AI model's predictions [10]. It is a strong indication that the AI model can provide trustworthy price forecasts. Stock trading relies on

identifying trends within a given period, and if the predicted prices align with the trend, they can be considered reliable for decision-making purposes.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

## **2. CHALLENGES**

In order to build the project, a few challenges have been identified as follows.

### **2.1. Getting information**

To predict prices with social media, I need to get the complete stock price and stock ticker from the website. It is hard to select the proper information because the stock has many different indicator prices for a single day. I need to only get the valuable number that can help the AI to predict the stock price with the tweet count. It is because the AI is more familiar with the number, I need to assign a number to each stock ticker in my model of prediction. This work can enable the AI to predict within a more precise way. After coding the model, I need this to be employed in the flutterflow, which enables the final stage of the model to be displayed in the app form.

### **2.2. Writing Paper**

It is hard to complete the research paper. For my project, I need to search for other popular methods of predicting prices and compare it with my project. I need to understand the difference and write the comparison. In addition, concluding different parts of the assignments is really hard, such as the challenge and solutions., to write enough words without misleading and misunderstanding different parts of the paper. Because it is a research paper I need to write clearly and logically. It is hard for me to write every part of the essay well. Fortunately, the essay has different sections that can help me to finish it more completely. Therefore, I can have less pressure for organizing works for the essay.

### **2.3. Choosing a Topic**

I did not have enough time to choose a topic, and there are many ideas to consider. I chose the topic of using social media to predict stock prices because I am really interested in that topic. Stock market is risky and can let people make a lot of money as much as people lose. There are some researches based on social media to predict, but people mostly analyze the correlation instead of the predicted price of the stock ticker. Researchers right now do more research on the sentiments about social media or the short-term influences about the stock price. My project only focuses on the tweet counts number. I believe that my project is a unique analysis aspect for the social stock topic. It will be easier for people who are not highly intelligent to have a reliable source for trading stocks.

### 3. SOLUTION

At first, we needed twitter and yahoo finance API to get stocks data and the tweet counts for the system.

```
def get_stock_tweets(ticker):
headers = {
    'Authorization':
    'Bearer
AAAAAAAAAAAAAAAAAAAAAAAAAMJulgEAAAAA7xfR9wRueDPmDSzSlcny1DJyQGM3DaeF3XqIVt
VIN7Cq0nd5GNbu9PCWGVIGF5K0KUw9wt7dvdn650S'
}
r = requests.get(
    f'https://api.twitter.com/2/tweets/counts/recent?query=
{ticker}&granularity=day',
    headers = headers)
r = r.json()['data']
for i in r:
    dict = i
    dict['start'] = dict['start'][0:10]
    y, m, d = dict['start'].split('-')
    dt = datetime.datetime(int(y), int(m), int(d))
    dict['start'] = dt.weekday()
    dict['DateInt'] = (dt.year * 10000 + dt.month * 100 + dt.day)
```

These codes can help us to retrieve the twitter information and format it. The first step, we got the tweet count for each stocks, and we will get the stocks data following the same process.

```
def update_info() :
    dt = datetime.now()
    x = dt.weekday()
    # print(x)
    if x != 6:
        return
list = [ ]
for ticker in stocks.keys():
    list.append(get_stock_info(ticker))
z = pd.concat(list)
z.to_csv('stock_master.csv', mode = 'a', index=False, header = False)
These codes can help us get data from existing stock tickers for both stock and tweet info and save them to csv files.
```

```
def get_stock_info(ticker):
info = yf.download(tickers=ticker,
    period = "1wk",
    interval = "1d",
    ignore_tz = True,
    prepost = False)
if(info.empty):
    raise Exception
info = info.drop(columns = ['Volume', 'Adj Close'])
info['ticker'] = ticker
These codes can help us to retrieve the stock data and format it.
def get_io_data( tweetCSV, stockCSV):
    a = pd.read_csv( tweetCSV )
    tList = [ ]
    for index, row in a.iterrows( ):
        if row['start'] l = 4 and row['start'] != 5:
            tList.append([row['start'], row['tweetcount'], row['ticker']])
```

```

b = pd.read_csv(stockCSV)
sList = [ ]
for index, row in b.iterrows( ):
    sList.append([row['Date'],row['Open'],row['ticker']])
return format_io(sList, tList)

```

These codes can help to read data and clean out the irrelevant data. This is because some information is unnecessary for AI to predict price. We only need to have an open price, highest price, lowest price and the closed price. These four information about the stock price will also appear on the app. After getting enough data and formatting it into csv files, we can make the data into the form that AI can read. It is because AI is more sensible with numbers compared with stock tickers. Therefore, we let each stock ticker name equal to a number starting from 1, so that AI can read the stock ticker, stock prices and the tweet count easily. The AI will use the past day information to find the correlation between stock price and the tweet counts for each stock which is in our system. So, if people type the stock ticker name that is not in our system, it will show the error signal. If people type the ticker name that is in our system, it will show the today's open price, closed price, highest price and the lowest price. It will also show tomorrow's predicted price. Thus, the AI can create a reliable predicted price for us to consider. These are the splash screen and the search page of my app.

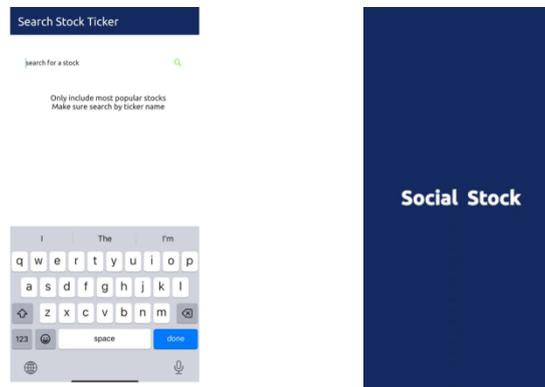


Figure 1. Screenshot of APP

## 4. EXPERIMENT

### 4.1. Experiment 1

I will do the experiment for three days. On each day, I will collect three stocks' predicted prices and compare them tomorrow. I will check five data points to compare the predicted data, such as the highest price, lowest price, open price and closed price and if it appears in the interval of the prices. But the final results will only record the difference between the predicted price and tomorrow's closed price. The sample is enough for practicing whether the app is able to provide reliable data because I will do this experiment 3 times. The result of the experiment will provide information about the differences with the tomorrow's closed price and the predicted price and the whether the price appears in the interval of the prices.

	D1 difference	D2 difference	D3 difference
AAPL	1	1.3	0.5
MSFT	0.9	0.7	0.4
GOOG	2	1.8	2.2

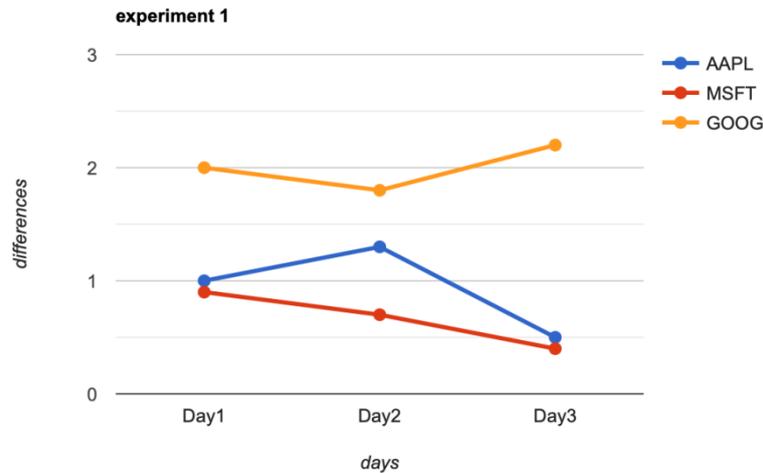


Figure 2. Figure of experiment 1

The differences between the real closed price and the predicted price are all smaller than 3, and AAPL and MSFT predicted prices appear in the real price's interval in those three days. Therefore, the experiment's results turn out that the app is reliable for people.

#### 4.2. Experiment 2

I will let four of my friends help me to do the experiment. Each of them needs to record the predicted price for 3 random stocks that are in the AI model and record the real price tomorrow for three days. They need to give the feedback of whether the price appears in the interval of the prices and calculate the differences between the predicted price and the real price at the end point of the stock market. Because the sample takes the random treatment, the sample is scientific and can provide reliable information. After that, I will make a conclusion of all results and identify the problems and provide a chart that concludes all experiment results.

	D1 difference	D2 difference	D3 difference
AAPL	1	1.3	0.5
MSFT	0.9	0.7	0.4
GOOG	2	1.8	2.2
TSLA	1.8	1.5	1.2

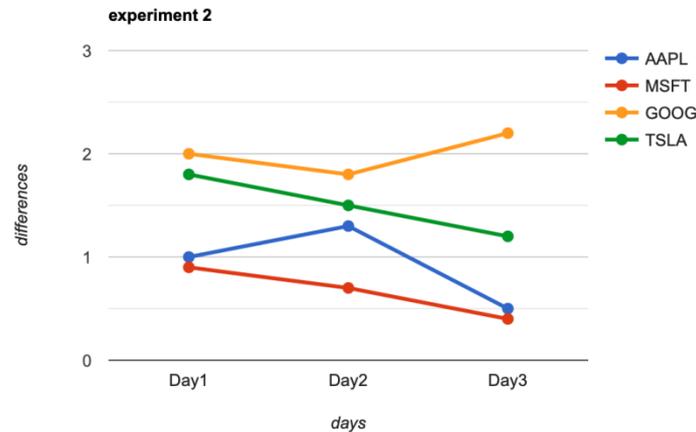


Figure 3. Figure of experiment 2

The predicted price between all stock tickers that is experimented in the experiment has the difference smaller than 3. And four of my friends indicated that tomorrow's predicted price appears for most of their stock tickers to be in the interval between the lowest price and the highest price.

According to all experiments we have done, the app can provide reliable data for people to consider. This is because the highest difference between the real price and the predicted price is less than 3, and all stocks in the experiments mostly appeared in tomorrow's real stock prices. It is because the stock market always changes when it is open. The indicator number for us to compare is not the every single price that appears in the stock market. The indicator number for our experiment to compare is the price at the endpoint in the stock market. Therefore, it is reasonable to understand that the AI model can not predict a price that is exactly the same as the real closed price. It is because the real stock price have only less than 3 differences with our system's predicted price. Thus, it is a strong evidence for indicating our AI model can predict the reliable prices. Trading stocks is more based on a trend in a period, so if the price we predicted in the trend, it is reliable.

## 5. RELATED WORK

The article Trading on Twitter: using social media Sentiment to Predict Stock Return is written by researchers Hong Kee Sul, Alan R.Dennis and Linyao Yuan [11]. Their findings are based on the fact that sentiments can spread really fast through social media, and specific companies with less than 171 followers will have a big impact. I only compare the tweet count of each company despite whether it is positive or negative, but these researchers analyze the emotion of the tweets users. I believe that their work can more precisely give the predicted stock returns than mine because they consider more factors than me.

The article Using News Articles to Predict Stock Price Movements is written by Győző Gidófalvi [12]. The researcher conducted the experiment by finding the relationship between the news article and stock price movement in the short run. The researcher found that even though the predictive power of the classifier is low, there is a strong correlation between the news articles and the behavior of stock prices from 20 minutes prior and to 20 minutes after news articles become publicly available.

The finding of the correlation is really valuable. My work is finding the correlation between the tweet counts and the stock price. My app can provide a specific number while the researcher's work can provide a correlation.

The article Predicting Stock Prices Using LSTM is written by Murtaza Roondiwala, Harshal Patel and Shraddha Varma [13]. They choose to use the LSMT to predict price because Long Short-Term memory is one of the most successful RNNs architectures. LSTM has the memory cells which enable the networks to effectively associate with memories. We have different prediction indicators, nevertheless we both get good results. Both strategies can give a reliable source for people who are trading in the stock market.

## 6. CONCLUSIONS

My app is based on using the ai model to predict stock prices. It got information like the stock price for yahoo finance API and the tweet counts from Twitter API [14]. After we get enough data, we use the data processor to format the data. Saving the data into the csv files and assigning the number for each stock ticker, which is easier for AI to read. Letting the data be a form that can be accepted by ai. Eliminating irrelevant data like some unnecessary stock price information that we download from the yahoo finance API for predicting stock price are also in progress when AI needs to predict the stock price. When we make the app in the flutter flow, we add different actions when people use the app. After typing the stock ticker name in the search bar, the app will show up the following information, such as the tweet counts, today's stock price, and tomorrow's predicted price. According to the experiment results, I can see that although my app can provide rationale results which are reliable, it can predict precise or accurate results. Although it is not exactly the same with the end point price. I think this problem occurs because the database is not big enough, so the AI cannot predict an accurate price.

The accuracy needs to improve because right now the AI can only provide a rational predicted price but is not really close to the real tomorrow price. According to the experiment that we have done, we find that although the price that AI predicts is logical and not ridiculous, it still needs more data on the stock price to help the AI model to find the relationship between each stock and its tomorrow price [15]. The practicability also needs to improve because it only has a few amounts of stock tickers in the system. Thus, it cannot predict every stock ticker in the stock market. The method of using social media to predict the stock price is helpful because the stock market is complicated, and additional predicted data can help people to make more money or lose less money.

Right now, I only use the tweet count to predict the stock price. In the future, I can get the information on more social media like reddit or the google trends. I can also improve the data processor and the stock pricing API.

**REFERENCES**

- [1] Preda, Alex. "Socio-technical agency in financial markets: The case of the stock ticker." *Social Studies of Science* 36.5 (2006): 753-782.
- [2] King, Benjamin F. "Market and industry factors in stock price behavior." *the Journal of Business* 39.1 (1966): 139-190.
- [3] Frino, Alex, and David Gallagher. "Tracking S&P 500 index funds." *Journal of portfolio Management* 28.1 (2001).
- [4] Bush, Robert R., and Frederick Mosteller. "A mathematical model for simple learning." *Psychological review* 58.5 (1951): 313.
- [5] Zhu, Ke. "Bootstrapping the portmanteau tests in weak auto-regressive moving average models." *Journal of the Royal Statistical Society: Series B: Statistical Methodology* (2016): 463-485.
- [6] Aggarwal, Rajesh K., and Guojun Wu. "Stock market manipulations." *The Journal of Business* 79.4 (2006): 1915-1953.
- [7] Piccolo, Domenico. "A distance measure for classifying ARIMA models." *Journal of time series analysis* 11.2 (1990): 153-164.
- [8] Chong, Terence Tai-Leung, and Wing-Kam Ng. "Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30." *Applied Economics Letters* 15.14 (2008): 1111-1114.
- [9] Gold, Steven. "The Viability of Six Popular Technical Analysis Trading Rules in Determining Effective Buy and Sell Signals: MACD, AROON, RSI, SO, OBV, and ADL." *Journal of Applied Financial Research* 2 (2015).
- [10] Wang, Zeyu, and Ravi S. Srinivasan. "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models." *Renewable and Sustainable Energy Reviews* 75 (2017): 796-808.
- [11] Sul, Hong Kee, Alan R. Dennis, and Lingyao Yuan. "Trading on twitter: Using social media sentiment to predict stock returns." *Decision Sciences* 48.3 (2017): 454-488.
- [12] Gidofalvi, Gyoza, and Charles Elkan. "Using news articles to predict stock price movements." Department of computer science and engineering, university of california, san diego 17 (2001).
- [13] Roondiwala, Murtaza, Harshal Patel, and Shraddha Varma. "Predicting stock prices using LSTM." *International Journal of Science and Research (IJSR)* 6.4 (2017): 1754-1756.
- [14] Bloch, Joshua. "How to design a good API and why it matters." Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications. 2006.
- [15] Zhu, Hao. "Big data and artificial intelligence modeling for drug discovery." *Annual review of pharmacology and toxicology* 60 (2020): 573-589.