Journal of Advanced Research in Business Law and Technology Management
Volume 3, Issue 1 - 2020, Pg. No. 10-14
Peer Reviewed Journal

Article

Sentiment Analysis used in Prediction of Stock Market

Shruti Saxena[1], Rahul Rajkumar Singh[2]

Research Scholar, MCA, Thakur Institute of Management Studies, Career Development and Research, Mumbai, India.

I N F O

Corresponding Author:
Rahul Rajkumar Singh, Research Scholar, MCA, Thakur Institute of Management Studies, Career Development and Research, Mumbai, India.
E-mail Id:
rahul567880@gmail.com
How to cite this article:
Saxena S, Rajkumar RS. Sentiment Analysis used in Prediction of Stock Market. *J Adv Res Busi Law Tech Mgmt* 2020; 3(1): 10-14.

Date of Submission: 2020-08-21
Date of Acceptance: 2020-08-30

A B S T R A C T

There are various techniques used by the investors for the prediction of future price of a stock or the trend of the entire stock market. Few parameters on which the movement of the stock depends are the performance of the company, political events such as the sudden fluctuation in the rise or fall of the currency, etc. This paper majorly focuses on the sentiments of the investors which depend on few parameters such as any scam news, current condition of the whole economy, peer organization execution, and so forth. In view of these boundaries, the supposition of a financial specialist might be positive, negative or unbiased. Here, we will perceive how the estimations of a speculator affect the cost of the stock so we can anticipate the right cost of the stock. Suppositions assume a significant job in the financial exchange and we can come to realize how financial exchange responds to various sort of information climate real or phony. Consequently, a methodology of assumption examination for securities exchange expectation has been arisen to foresee the effect on financial exchange dependent on human opinions.

Keywords: Stock Market, Stock Price, Sentiments

Introduction

Stock Market or Share Market is where the purchasers and dealers can purchase and sell the protections of those organizations that are recorded on a stock trade. A huge organization generally has its stocks recorded on many stock trades the world over. There are different stock merchants and electronic exchanging stages where the exchanging of stock happens. The interest in financial exchange is done dependent on some speculation technique as a primary concern. An endeavor which is made to foresee the future cost or pattern of the stock is known as financial exchange expectation. This prediction is made using various future price prediction methods. The future value of a stock obtained from any of these methods must be robust and accurate because the investors may lose a huge amount of their hard earned money if this prediction goes wrong. All the real world scenarios and the other variables that will affect the price of the stock must be taken into the consideration while implementing the prediction system. Many scientific attempts have been made to correctly predict the movement of a stock but no perfect method has been derived yet which will generate the accurate result. While designing a model for prediction, there are basically two parameters which need to be considered: Technical and Fundamental approaches. In addition to this, the human sentiments also play an important role in the prediction of stock prices. In this paper, we focus on considering the human sentiments that how it is related to the movement of a stock price. Sentiment is considered as a judging opinion as positive, negative or neutral. On

the basis of this, experts make a prediction for the stock as buy, hold or sell. These predictions are been made based on the news that are related to a particular company. It may be positive or negative news for that company. When there is any negative news related to the company, there is a negative impact in the minds of the investors which eventually has an effect on the price trend of the stock. A detailed description of the relation of the human sentiments to the trend of the stock is explained in this paper.[1]

Literature Survey

The future price of a stock or the trend of the stock market can be predicted by using the following techniques:[2]

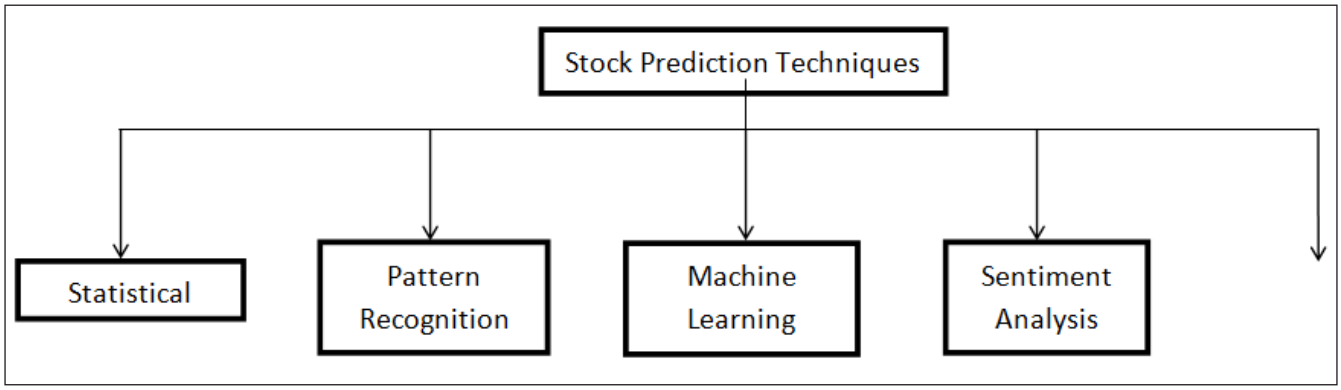

Figure 1

The paper describes the stock price prediction by making use of sentiment analysis by understanding the sentiments of an investor. The prediction of future price of a stock is actually a tough task. From the past few years, many changes have been taken place in the finance sector. The market experts are behind the search of new methods which will predict the price of the stock more accurately. Various researches are been made for the purpose of price prediction.

Anass Nahil[1] proposed a method which was newly developed for stock price prediction. It is used by many investors and stock market experts to gain as maximum benefits as possible. This system which was proposed by Anass was based on Support Vector Machine (SVM). It was largely based on discriminative classifier algorithm which gives an accurate data. This experiment was performed on stocks of 5 banks. The conclusion of this experiment was that the SVM performs in a more optimised way when the global evolution of all the market is added to the independent variables. Thus the three indices of Casablanca Stock Exchange are used which are MASI, MADEX and Banks Sector Index.

Narendra Pahuja, Abhishek Oturkar, Kailash Sharma, Jatin Shrivastava, Dimple Bohra[2] used a model named ARIMA Model which was much good predictor. They observed that the stock market is non-linear, dynamic and chaotic in nature. They used the data from the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) for the prediction of stock price. From the observation of this experiment, it was observed that the ARIMA model proved to be quite effective with accurate price prediction as compared to the existing prediction model.

NP Samarth, Gowtham V Bhat, Hema N[3] made a research on the stock price prediction. This prediction was made using some of the machine learning algorithms and the future stock price of a company was predicted. Since the time series and the fundamental and technical analysis are very unreliable and limited, this paper was proposed. Python was used for the implementation purpose. They proposed an approach that used machine learning algorithms and training will be made on the historical stock data that is available and gain intelligence. Later it used the knowledge acquired for predicting the stock prices accurately. Random Forest Regression was used for stock price prediction for small and large capitalizations also in different markets employing both up-to-minute and daily frequencies.

All those above mentioned papers considered various parameters for making a prediction but they didn't consider the human sentiments into picture.

Relevance of the Project

There are many people in this world who have very less idea regarding the investment strategies in the stock market. Some of them just blindly invest their money in any of the stock without thinking about anything. Later on, there are high chances that the investors will lose their money because of such blind investment. Various parameters play an important role while taking the decision whether to invest in a particular stock or not and human sentiment analysis is one of them. This research is done on the trading volume of a particular stock and its direct relation with its price. This will, to a certain extent, help the investors to invest their money in the right stock and prevent themselves from losing their capital money.

Proposed Methodology

Data Collection

The sentiments of an individual are based on the news he receive from various sources. These sources might be through internet, economic TV channels, newspapers etc. Various API's of certain finance related applications such as Economic Times can be used to fetch the news but it's quite complicated to get those API's on the internet for free of cost. So, another alternative for generating the dataset is that we can create a CSV file which will contain some news related to some of the companies which have their shares listed on the Indian stock market. This CSV file needs to be generated manually in which the data is collected from various sources. This file will be just used for representational purpose.

Feature Extraction Module

The dataset which is generated manually needs to be worked on. The news which is present in the dataset is to be cleaned and the keywords need to be extracted from it. For this purpose, we have a predefined set of dictionary words which are present in a CSV file. This dictionary is used to compare the words present in our dataset and the keywords are extracted from the news description. Based on the dictionary, the words are classified as positive words and negative words. Accordingly, two different word clouds are formed: Positive word cloud and Negative word cloud. For example, the word huge profits will be included in the positive word cloud and the word lower circuit will be included in the negative word cloud.

Training Module

For the purpose of creating a strong dataset, we need to train the keyword extraction module. There are many things present in a sentence such as punctuation marks or other things like @, $, #, etc. Those are known as stop words. These stop words are not of any use and hence they must not be included in our dataset unnecessarily. For this purpose, we have to train our model so that only the relevant words are extracted from the sentence and other stop words are eliminated from getting stored in the dataset.

Prediction Module

Based on the positive word cloud and negative word cloud, the sentiments of a human can be predicted. The dataset which is used contains the recent news of some companies. Based on the recent news, the keywords are extracted, classified and stored in their appropriate cloud. For representational purpose, we have considered the news and reviews of a single company, Reliance Industries. Once the word cloud is generated, we have to take a count of the number of words present in each cloud. If the number of words in the positive word cloud is more than the number of words in the negative word cloud, we can say that the trend of the stock is positive because there's some positive news associated with that particular company from the past few daysdays.

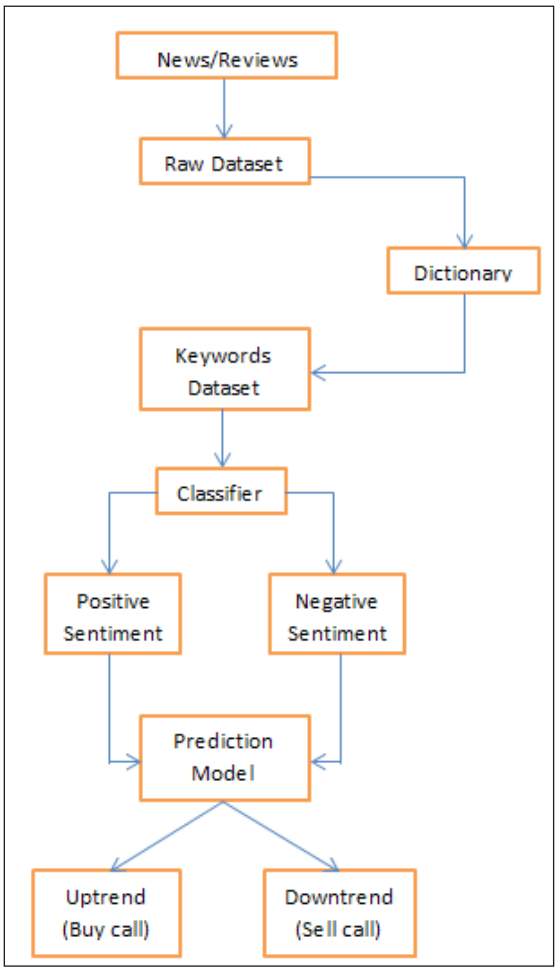

Figure 2.Work Flow

Experimental Study

First, we created a dataset. In that dataset, we added the reviews and news regarding a particular company. The reviews or news which we included in the dataset are the most recent ones which are about a week or a month old. We will store all the text reviews in a comma separate value i.e. CSV form by having only one attribute in that CSV file name as Headline. So, while implementing the task, firstly we will read the CSV file in our jupyter notebook using pandas library. After reading the CSV file we will show only first 5 review data using head command. After reading the CSV file we will describe the CSV file using review csv describe command.It will give us the total count of words, unique words, top and frequent words.

Later, we will create one function that will give us wordcloud using some stopwords that can be present in the file. After

that we will do some statistical analysis on the basis of wordcount, charcount worddensity and punctuation density using lambda, split and replace function. And will show as output by showing the head of that statistical analysis data and then describe that statistical data.

After this we will create two functions which will give us the polarity and subjectivity of the words present in the file and will add this polarity and subjectivity in two data frames name as df1 and df2. On the basis of that dataframes, will find the range for that polarity and subjectivity and will join with the earlier output.

When we join the dataframes, we will plot the word cloud of most frequent words used in our dataset. After getting the most frequent words we will find the positive wordcloud and also find the count of the positive wordcloud. We get positive wordcloud when the polarity range is greater than zero.

```
In [21]: review_csv.describe()
```

Out[21]:

	Headline
count	33
unique	32
top	The consolidated profit in the previous quarte...
freq	2

Figure 3.Describing the CSV File

```
28]: print(df1.head())
     print(df2.head())
```

```
   polarity
0 -0.028571
1  0.000000
2  0.066667
3  0.197778
4 -0.155556
   subjectivity
0      0.483333
1      0.300000
2      0.466667
3      0.757778
4      0.288889
```

Figure 5.Describing Subjectivity and Polarity

Out[23]:

	Headline	word_count	char_count	word_density	punc_count
0	Company's industries share price rallies 10% o...	45	240	0.186722	5
1	Company's capital share price falls 5% hits 52...	28	169	0.164706	5
2	Company's communication hits lower circuit for...	41	215	0.189815	5
3	Company's hits 52-week high on strong Dec quar...	31	145	0.212329	3
4	Company's Industries stock price slips 4% in i...	25	131	0.189394	8

```
In [24]: #Statistical analysis
         review_csv.describe()
```

Out[24]:

	word_count	char_count	word_density	punc_count
count	33.000000	33.000000	33.000000	33.000000
mean	37.363636	189.606061	0.197377	5.272727
std	12.948982	67.448100	0.019914	3.947093
min	12.000000	65.000000	0.153846	1.000000
25%	27.000000	145.000000	0.186722	2.000000
50%	38.000000	197.000000	0.198653	4.000000
75%	45.000000	240.000000	0.206897	7.000000
max	59.000000	312.000000	0.257426	20.000000

Figure 4.Statistical Analysis

```
print("Total Number of Positive words = ", total_positive)
```

```
Total Number of Positive words =  129
```

Figure 6.Positive Word Count

After the positive wordcloud we will now find the negative wordcloud and also the count of the negative words present in the file. We get the negative wordcloud when the polarity range is greater than zero.

```
print("Total Number of negative words = ", total_negative)
```

```
Total Number of negative words =  124
```

Figure 7.Negative Word Count

After getting the wordcloud we will know where our trend lies, whether it is positive, negative or neutral. So, if we get the trend as positive we will suggest buyer to buy the stock and if we get trend as negative we will suggest buyer to opt out from this stock and on the basis of neutral that is positive and negative are same it will be his own decision whether to buy this stock or to opt out.

```
if(total_positive>total_negative):
    print("Trend is showing towards positive you can buy the stock.")
elif(total_positive<total_negative):
    print("Trend is showing towards negative it will be risky to buy stoc
else:
    print("Trend is neutral you can go with your decision.")
```

```
Trend is showing towards positive you can buy the stock.
```

Figure 8.Trend Analysis

At the end we will give the correlation graph on the basis of polarity and subjectivity ranges.

Out[35]:

	word_count	char_count	word_density	punc_count	polarity	subjectivity
word_count	1.000000	0.973753	0.013342	0.567228	-0.173049	0.048099
char_count	0.973753	1.000000	-0.197848	0.609983	-0.199703	0.013635
word_density	0.013342	-0.197848	1.000000	-0.202912	0.157086	-0.008625
punc_count	0.567228	0.609983	-0.202912	1.000000	-0.178456	-0.140525
polarity	-0.173049	-0.199703	0.157086	-0.178456	1.000000	0.149397
subjectivity	0.048099	0.013635	-0.008625	-0.140525	0.149397	1.000000

In [36]:
```
#let's visualize polarity vs subjectivity
plt.scatter(review_csv.polarity, review_csv.subjectivity, color='r')
plt.show()
#so...many reviews are actually in (kinda) neutral region
```

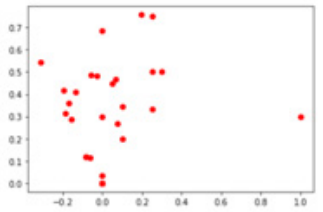

Figure 9.Correlation Graph

Result Analysis

So, the result analysis we get from this experiment. As there are many new buyers of stock. And many don't know how to invest or on what basis we have to invest. So, on the basis of our Stock review sentiment analysis we will show buyers some positive and negative count and where the trend lies. And also suggest them, what to do with related stock whether to buy or to opt out. Our result analysis mostly depends on the wordcloud and the word count of the negative, positive reviews by identifying the polarity and subjectivity of the words and ranging them accordingly by using some python libraries. So, the records present in the file will perform some data science and will the give the analysis regarding the stock of that company. So here, when we performed the experiments on the dataset, we found that it contains 129 positive words and 124 negative words. It is observed that the positive words are more than the negative words which means there are many positive things happening within the company. It will have a positive impact on the stock of that company in the mere future. So, based on this analysis, a buy call can be given to the invester.

Conclusion

Indian stock market is full of ups and downs. Hence it is quite important that an investor must invest the money wisely after studying the fundamentals of a company. Based on the experiments we will perform, we are expecting a model that will predict the trend of the stock. Based on human sentiments i.e. the comments made by the individuals of stock experts, a dataset will be formed and a word cloud will be generated. If the size of the positive word cloud is greater than the size of the negative word cloud, buy call can be given or vice versa.

References

1. Anass N, Lyhyaoui A. Stock price prediction based on SVM: The impact of the stock market indices on the model performance.
2. Pahuja N, Oturkar A, Sharma K et al. Stock Market Prediction using the ARIMA Model 3.
3. Samarth NP, Bhat GV, Hema N. Stock Price Prediction. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, 2019; 9(2).