Brainwave: A Multidisciplinary Journal, Vol. 3, No. 2, June 2022, pp. 137-143, © Brainware University

Effect of Epidemic in Stock Market Concern and GDP

Sandipan Biswas*
Department of Computer Science and Engineering,
Brainware University, West Bengal, INDIA
sandipan diet@rediffmail.com,

Shivnath Ghosh
Department of Computer Science and Engineering,
Brainware University, West Bengal, INDIA
dsg.cs@brainwareuniversity.ac.in,

Sandip Roy
Department of Computational Science,
Brainware University, West Bengal, INDIA
sandiproy86@gmail.com

**Corresponding Author*

Abstract

In the history of the world, no illness as virulent because the Novel Corona Virus (COVID-19) has forced international panic. As the threatening illness, international economies and stock markets are witnessing uncertainty in trade. Researchers use sentiment analysis to sight prevailing sentiments of users from posts and comments denote on social media on the web. In this paper, we have proposed a technique to predict the capitalist reaction to news and developments associated with the Corona Virus. We have used Python language and TextBlob, a Python library, to demonstrate how our planned model will forecast market trends when aggregating tweets regarding the Corona virus pandemic. We have also additionally discussed the literature, involving sentiment analysis and language process techniques which are utilized by researchers for prediction models. Supported information extracted from social media like Twitter. In our analysis experiments, we have also used extract news, numerical information from Yahoo!, Finance and Twitter, to make an instrument and prediction model. Results of our experiments reveal that our planned model was ready to market trends, and also the forecasts corresponded to actual movements of the metropolis securities market index, SENSEX.

Keywords*: COVID-19; Sentiment Analysis; Stock Market; TextBlob Algorithm.*

Introduction

The contemporary global spread of the Corona virus [1–3] has led to a significant interruption in international supply chains. The Baltic Dry Index, measuring demand for international shipping capacity, lost 45% between January 02 and February 28 2020, reflecting a decrease in aggregate economic production and supply. This reduction has fueled a historic loss for Dow Jones (the dominant US stock market index) of 12.4% in just one week between February 21 and February 28 ,2020. Besides the interruption of economic supply flows, a major economic worry is that spreading fear of the Corona virus will weaken economic sentiment. As economic research documents, individually form their macroeconomic expectations from current events, news and experiences [4–6]. A spreading fear of the Corona virus might therefore hamper economic expectations in the current and future state of the economy. The Director of the World Health Organization, Tedros Adhanom Ghebreyesus, warned precisely of such a plausibility on February 28 when he stated that "stigma[...] is more dangerous than the virus itself. Fear and panic are dangerous". Canonical theories of economic demand and the psychology of markets [7–9] as well as his evidence which are [10] highlight detrimental effect of economic expectations on economic demand, which, if large enough, has the potential to trigger a recession. Therefore, using two complementary methodologies, we assess the impact of fear of the Corona virus on economic sentiment, and the policy implications for the prevention of an upcoming recession. Firstly, we collected global data on the intensity of Internet searches from Google Trends to measure economic fears. As shown by prior studies, such Internet searches are accurate predictors of future economic demand and activity as these capture the sentiment on the consumer's side of the economy [11,12]. We validated this claim by relating economic output as well as individual components of aggregate demand to the pre-quarter search intensity for the Google search topic 'Recession' in country-level regressions controlling for country and year-by-quarter fixed effects.

Using quarterly data from 2015 to 2019, we find that real GDP growth and real growth in consumption and imports are significantly lower in the quarters following increases in recession topic searches (Fig1A). An 100% increase in search intensity for recession related topics is associated with a 1.6 %-point lower consumption growth rate and a 1 %-point lower GDP growth rate in the following quarter. Hence,these search intensities are a leading indicator of subsequent aggregate demand contractions and economic downturn.

Related Work

For mining the data, there are different text mining approaches which can be used.

Dhirajgurkhe, Niraj pal and Rishit Bhatia proposed how Twitter data is processed. Firstly, they collected data from various sources and eliminated those features which do not contribute to any polarity. Then, this data is sent into the sentiment classification engine, i.e., naïve bayes classification algorithm which calculates the probabilities, which is to say how much data is corrected, and predicts the sentiment for the given query.[7]

M.Bouazizi, T. Ohtsuki have proposed that tweets which contain more than one sentiment be called multi class sentiment analysis. They have identified the exact sentiment conveyed by the user rather than the whole sentiment of the tweet. They have also used the SENTA tool to identify this thing. They proposed an approach with the help of which they have calculated the sentiment score. Sentiment is having highest score that will be considered this process is called as "Quantification". [8]

Geetikagautam, Divakar Yadav have discussed customer review classification for which they have used already labelled Twitter data set. For the purposes of this task, they have used machine learning based algorithm i.e. Naïve Bayes, SVM, Maximum Entropy. They have worked on Python and NLTK for training the SVM, Naïve Bayes, Maximum Entropy. Naïve Bayes is better technique in term of accuracy and gives better results when compared to Maximum Entropy. We can get better results by comparing to SVM by using the SVM with the Unigram model and then further accuracy can be improved by semantic analytic, followed by wordNet.[9].

Akshay Amolik, Niketanjivane, Mahavir Bhandari and Dr. M. Venkatesan have discussed a highly suitable model in their paper which will take the twitter data of upcoming Hollywood and Bollywood movies. They are able to do this task with the help of a classifier, and features like SVM and Naïve Bayes. Both of these are used for their high accuracy, but in terms of precision, Naïve Bayes is better than SVM, whilst SVM is better than Naïve Bayes in terms of its recall. By increasing the dataset, we can increase the classification accuracy.[10]

Subhabrata Mukherjee et al. have discussed a hybrid system named as TwiSent which has resolved problems like spam tweet, pragmatics and noisy texts. Twisent consists of a spell checker and a pragmatics handler. The Spell checker detects the noisy text, whereas the pragmatics handler handles the pragmatics in tweets.Twisent yields better results compared to C-feel-IT system. The accuracy of finding the negative sentiment of Twisent system is higher than C-Feel-IT. [11].

Dmitry Davidov, Oren Tsur and Ari Rappoport in their paper, have proposed a supervised sentiment classification structure which works well with Twitter data. They have used the K-nearest neighbor and featurevector. The basic purpose of this framework is to identify and distinguish between the sentiment types defined by smileys and tags.[12]

Neethu M. S. and Rajasree R have used the machine learning techniques in their survey paper to explore the Twitter data related to electronic products. They have used a feature vector for the tweets classification. They have used three types of classifiers, i.e. SVM, Naïve Bayes, and Maximum Entropy, and these classifiers were tested using the Matlab simulator. SVM and Naïve Bayes classifier are implemented using a built-in function, whereas the MaxEnt classifier is used by MaxEnt software. So basically, the all classifier has nearly the same performance[13].

Pulkit et al. built and proposed a model which extracts tweets from Twitter based on post-terror activities. They made their study on the terrorist attack which occurred in URI on 18th September, 2016. They studied 59,988 tweets issued in the aftermath of the attack, while only considering tweets with the hashtags #UriAttack, #uriattack, #uriattacks. They have used the Naïve Bayes and SVM to extract the time of the last re-tweet and the number of retweets [14]

Sudarshan Sirsat et al. proposed a technique in sentiment analysis on Twitter data where they collected reviews of the product. They have used the Naïve Bayes algorithm which performs better in term of accuracy and efficiency. They have extracted 200 tweets with an average length of 70.105 characters. The aim of this research is to identify the characteristic of tweet like how many times the tweets were liked and how many times they have re-tweet the tweet.[17]

Hetu et al. built and proposed a model in sentiment analysis on Twitter data based on Anaconda Python. They have extracted the dataset from Kaggle , which they have classified the people emotions based on positive and negative reviews. This model gives high accuracy for large datasets.[16]

Ali Hasan et al. proposed a model using the hybrid approach that comprises of Sentiment Analyzer Machine Learning. They took only those tweets that were followed by the hashtag (#) and current political trends. Basically, this model converts the Urdu tweets into English tweets. They took 1690 tweets for training data and 400 for testing the data. They used the Naïve Bayes and SVM Classifier for training the dataset in building a model. They used three different libraries to calculate the subjectivity and polarity.[15]

Feddah Alhumaidi Al Otaibi, et al. proposed a model by using supervised and unsupervised algorithms. They deployed sentiment analysis to determine which of the two restaurants, McDonald and KFC, is more popular. Moreover, they extracted 7000 tweets of both restaurants by Twitter API. The tweets were in English and they used R programming language. R programming language can perform huge computational tasks. They have used several machine learning techniques, but they found MaxEnt performing better than the other techniques. Moreover, they found KFC having many neutral tweets and McDonald having more positive and negative tweets as well.[14].

Proposed Model

After a day of volatile trade, the S&P BSE Sensex lost 581 points (or 2%) to 28,288 down on 19.03.20 (Thursday), its lowest closing level since February 2017. Fears of a global recession mounted despite massive stimulus measures from central banks and governments around the world. The broader Nifty 50 index slid down by 205 points or 2.4% to end at 8,263.

Proposed model component description

```
Input: Text File (News Headlines which include Nouns, Adjectives, Adverbs)
Output: Values > 0 (Positive), Values < 0 (Negative), Values = 0 (Neutral)
Begin:
    Sentiment Analysis () ← File
    For each row in rows
      if Sentiment Polarity Score(line) > 0 then
        Sentiment ← Positive
      else
        if Sentiment Polarity Score (line) < 0 then
          Sentiment ← Negative
        else
          if Sentiment Polarity Score(line) = 0 then
            Sentiment ← Neutral
          else
        end
      end
    end
end
```

Algorithm 1: Sentiment Classification using Text Blob

Here in Algorithm 1, we have represented algorithm details of our proposed model.

Sentiment Analysis Component:

In this component, the partiality of stock news data has been performed thus: For news headlines, the objective is to classify news in accordance with whether they bear either positive or negative sentiments. For achieving this, data pre processing is performed on news headlines, followed by news classification using the Naïve Bayes algorithm. The section underneath describes in details the steps of the proposed model.

Text preprocessing: There are several pre processing steps which are performed thus:

Tokenization: Each news headlines or financial report is split into meaningful words called tokens.

Data standardization: In this technique for data consistency, all words in articles and reports about companies are transformed in a document into lower case.

Stop-word-removal: Words which do not have significant meaning in a sentence such as “the, a, of…”etc., are removed to reduce the number of features and also enhance the reading appeal.

Stemming: Here, Porter Stemmer is applied on the data set to return each word to its stem and remove suffixes (such as -ed, -ing, -ion etc.) to reduce the complexity of the document and also minimize the processing time and improve the model performance.

Programming Language and Simulator

Python

Python is an interpreted, high-level, general-purpose programming language created by Guido Van Rossum in 1991, which highlights code readability, and is known for its notable use of significant white space. Thisconstructive and object-oriented approach aims at and helps programmers in writing clear, logical code for small and large-scale projects.

Anaconda Navigator

Anaconda is a free and open-source distribution for Python and R programming languages which is used in scientific computing for data science, machine learning applications, large-scale data processing, predictive analytics etc. The aim is to simplify package management and deployment.

Package versions are managed in the Package Management System Conda This Anaconda distribution includes data-science packages which are suitable for Windows, Linux, and MacOS. Anaconda distribution comes with more than 1,500 packages, as well as the Conda package and a virtual environment manager. It also includes a GUI called Anaconda Navigator with graphical alternative in Command Line Interface (CLI).

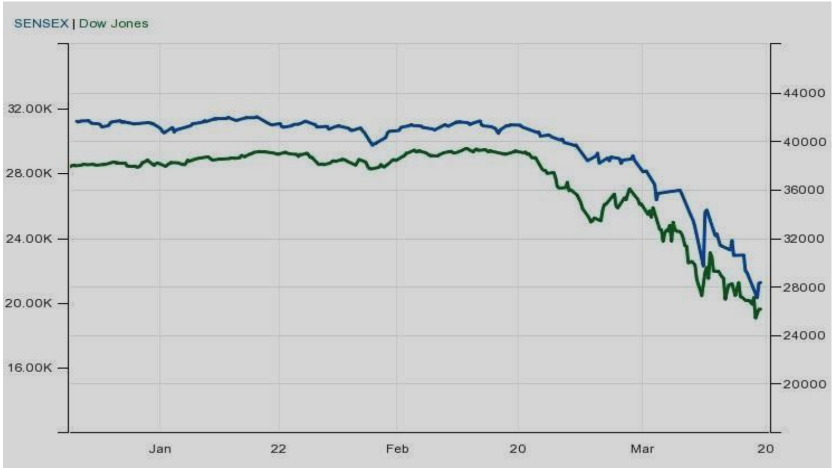

Fig. 1.Comparison of share of Sensex to Dow Jones (Yahoo Finance)

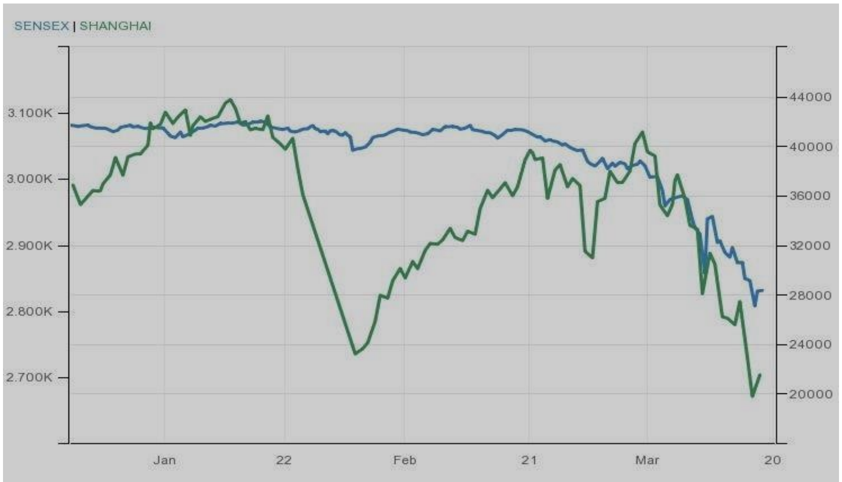

Fig. 2. Comparison of share of Sensex to Shanghai (Yahoo Finance)

Result Analysis

Here all the results are calculated on the database of the share market. The share market is a field where investors are affected by the views of traders and experienced people in the market. In the case of our model, few stocks have been taken and the results are then generated on the basis of the comments/reviews given by people on those stock items. If the review is positive, then it is supposed to buy or sell share. Here we have not looked at a particular share, but rather considered the overall stock exchange rate up and down.

Fig. 3. Sensex trend over a specific time period (Image generated from data sheet plotting to excel)

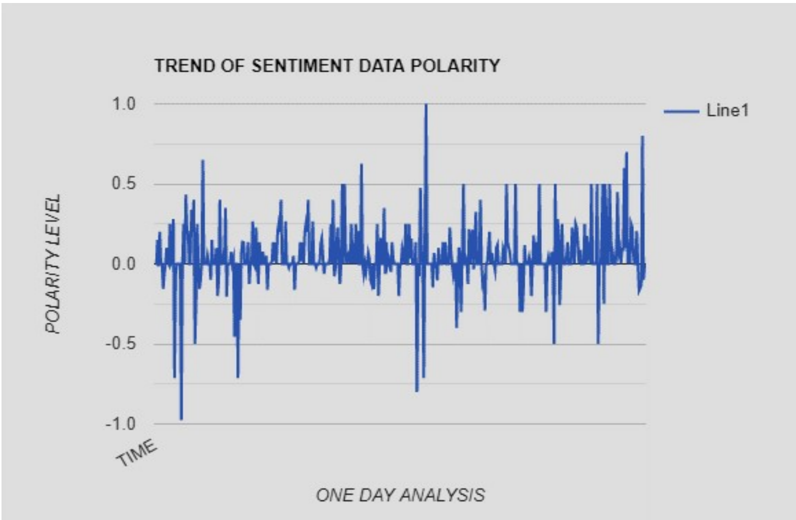

Fig. 4. Trend of sentiment data polarity (Image generated from data sheet plotting to excel)

Conclusion

In our planned model we have investigated on the concurrent result of analyzing different sorts of news with historical numeric values for understanding securities market trend. The planned model has worked onthe views/ opinions of the reviewers on the shares The views of the specialists have an effect on the traders to take a position into the market. This comparative study have proved the it. Our planned model has improved the prediction accuracy for analyzing the trend of the share market, by analyzing differing types of daily news with the help of entirely different values of numeric attributes throughout a time domain.

Reference

1. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Tan, W., & China Novel Coronavirus Investigating and Research Team (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England journal of medicine, 382*(8), 727–733.

2. Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet (London, England), 395(10225), 689–697.

3. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K., Lau, E., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., … Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine, 382*(13), 1199–1207.

4. Malmendier, Ulrike Nagel, Stefan (2011). Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics, 126*(1),373–416.

5. Coibion, Olivier, Gorodnichenko, Yuriy, Kamdar, Rupal. (2018). The formation of expectations, inflation, and thePhillips curve. *Journal of Economic Literature,56*(4),1447–91.

6. Carroll, Christopher. (2003). Macroeconomic expectations of households and professional forecasters. *The QuarterlyJournal of Economics, 118*(1),269–298.

7. Keynes, John Maynard (2018). *The general theory of employment, interest, and money*. Springer.

8. Akerlof, George A., Shiller, Robert J. (2010) . *Animal spirits: How human psychology drives the economy, and whyit matters for global capitalism*. Princeton university press.

9. Robert J Shiller.(2015). *Irrational exuberance: Revised and expanded third edition*. Princeton university press.

10. Matsusaka, John G., Sbordone, Argia M . (1995). Consumer confidence and economic fluctuations. *Economic Inquiry, 33*(2),296–318.

11. Vosen, Simeon , Schmidt, Torsten.(2011). Forecasting private consumption: survey-based indicators vs. googletrends. *Journal of Forecasting, 30*(6):565–578.

12. Choi , Hyunyoung, Varian, Hal. (2012). Predicting the present with google trends. *Economic record, 88,* 2–9.

13. Armona, Luis, Fuster, Andreas, Basit Zafar. (2019). Home price expectations and behavior: Evidence from a randomized information experiment. *Review of Economic Studies, 86*(4) .

14.Coibion, Olivier, Gorodnichenko, Yuriy, Ropele, Tiziano. (2020). Inflation expectations and firm decisions: Newcausal evidence. *The Quarterly Journal of Economics, 135*(1),165–219.