Macrothink Institute™

Business and Economic Research
ISSN 2162-4860
2015, Vol. 5, No. 2

Predicting Stock Market Indexes, Two Days In Advance, Using NewsInn Algorithm

Radu Nicoara

Dept. of Computer Science, Spiru Haret University, Bucharest, Romania

Tel: 4-074-851-2219 E-mail: nicoaraeradu@yahoo.com

Received: October 25, 2015 Accepted: November 15, 2015

doi:10.5296/ber.v5i2.8664 URL: http://dx.doi.org/10.5296/ber.v5i2.8664

Abstract

NewsInn is an A.I. Driven Algorithm that processes and conglomerates news from major news publications. It uses an opinion extraction algorithm to do a sentiment analysis on every news article.

Considering that stock markets are heavily influenced be world news, we conducted a study to show the link between the detected sentiment inside the news, and the most used Stock Market Indexes: S&P 500, Dow Jones and NASDAQ. Results showed an almost 70.00% accuracy in predicting market fluctuation two days in advance.

Keywords: Artificial Intelligence, Stock Market Index, Sentiment Analysis

1. About Newsinn

Description

NewsInn is an Artificial Intelligence Algorithm that parses the major news publications for news articles, groups them based on the main story, and presents them in order of importance. [1] It was design to become a news aggregator at first, but since important stories are covered be most publications, an Artificial Intelligence algorithm was implemented in order to group such articles.

Opinion Extraction

To do a sentiment analysis on each article, the algorithm used a derivative of SentiWord [2], but it quickly became clear that the library was not created to be used in a news Environment. So a new library was developed, that would better suit the case it was being used in. It manages to provide a good estimate of the impact that the text of an article has (see image 1).

Image 1. An example of an article's Sentiment Analysis Result

2. About Stock Indexes

Definition

A stock index or stock market index is a measurement of the value of a section of the stock market. It is computed from the prices of selected stocks (typically a weighted average). It is a tool used by investors and financial managers to describe the market, and to compare the return on specific investments. [3] [5]

Theory

Since the stock market is heavily affected by world problems, and therefore, world news, it is highly likely that a general sentiment analysis of the news would produce a result directly related to the general sentiment of the public on the world affairs. This would theoretically, in time, affect the stock market. The time-to market of world news we estimated to be of a few days, since only huge golbal events have the power to change the stock price in-day.

Therefore we started testing at a one-day offeset, then went on for two, three, and even four days offset between the news data and the stock data. We only kept the data that provided the highest accuracy in predicting the stock market indexes flow.

3. Data Collection

Data Aquisition

Taking into account the fact that the final version of the NewsInn A.I. Algorithm started gathering information, on the 14th of September, the NewsInn algorithm checked more than *24,000* articles against the Opinion Extraction library. We used this data to gather an average opinion score of the all the news written that day.

The opinion, however, was extracted from the conglomerated articles. We found *no resemblance* between the raw article opinion and the indexes. The explanation relies probably in the idea that people only register one event, even if, in some cases, it is detailed in more than 20 articles.

The Data extracted can be found below. Please note that week-end data was removed, as it is

not highly accurate (contains less than half the number of the articles that a week day has), and the fact that Stock Markets are closed during week-ends.

Table 1. Result of the Average Opinion of Conglomerated Articles (excluding week-ends)

Date	Opinion Score
14.09.2015	0,64
15.09.2015	0,61
16.09.2015	0,63
17.09.2015	0,60
18.09.2015	0,54
21.09.2015	0,53
22.09.2015	0,61
23.09.2015	0,58
24.09.2015	0,67
25.09.2015	0,48
28.09.2015	0,42
29.09.2015	0,53
30.09.2015	0,37
01.10.2015	0,59

Date	Opinion Score
02.10.2015	0,46
05.10.2015	0,18
06.10.2015	0,34
07.10.2015	0,44
08.10.2015	0,40
09.10.2015	0,49
12.10.2015	0,22
13.10.2015	0,37
14.10.2015	0,37
15.10.2015	0,50
16.10.2015	0,47
19.10.2015	0,21
20.10.2015	0,24
21.10.2015	0,32

We used the historical prices of ***Dow Jones Index***, the ***S&P 500 Index*** and the ***NASDAQ Index*** for analysis inside of the graphs, while offsetting them by ***two days*** in the case of Dow Jones and NASDAQ, and ***one day*** for S&P 500, to accommodate the prediction time of the algorithm. [4]

Table 2. Table used to plot the graphs and extract data

We arrived at the following data, normalized for ease of use:

Date	(Opinion+1.1)	((DJ/1000)-12)/2	((NASDAQ-1000)/400-6,3)	(S&P/1000)
14.09.2015	1,74	2,34	3,42	2,16
15.09.2015	1,71	2,19	3,43	2,19
16.09.2015	1,73	2,26	3,27	2,18
17.09.2015	1,70	2,17	3,27	2,12
18.09.2015	1,64	2,14	3,09	2,13
21.09.2015	1,63	2,10	3,08	2,09
22.09.2015	1,71	2,16	3,04	2,08
23.09.2015	1,68	2,00	2,92	2,06
24.09.2015	1,77	2,02	2,56	2,07
25.09.2015	1,58	2,14	2,49	1,96
28.09.2015	1,52	2,14	2,75	1,97

29.09.2015	1,63	2,24	2,77	2,04
30.09.2015	1,47	2,39	2,97	2,05
01.10.2015	1,69	2,40	3,15	2,10
02.10.2015	1,56	2,46	3,07	2,17
05.10.2015	1,28	2,53	3,18	2,16
06.10.2015	1,44	2,54	3,23	2,19
07.10.2015	1,54	2,57	3,28	2,23
08.10.2015	1,50	2,54	3,30	2,23
09.10.2015	1,59	2,46	3,19	2,23
12.10.2015	1,32	2,57	3,16	2,21
13.10.2015	1,47	2,61	3,38	2,19
14.10.2015	1,47	2,62	3,42	2,25
15.10.2015	1,60	2,61	3,46	2,27
16.10.2015	1,57	2,58	3,40	2,27
19.10.2015	1,31	2,74	3,30	2,26
20.10.2015	1,34	2,74	3,50	2,24
21.10.2015	1,42	2,74	3,50	2,31

Were data could not be found, we duplicated the last known data. That is, the last row of the table. But this row was not taken into account when doing the calculations. Please see below in blue.

4. Results

Dow Jones Index

The prediction, *two days in advance*, for the Dow Jones Index had the highest degree of accuracy, with an average prediction rate of 69.44%. For 4 of the days, for unknown reasons, the prediction was offset by a day, so it could only predict the moving of the index in a one day interval. This was probably due to the refugee crisis.

For the offestted case we divided the number of days correctly predicted by a factor of two.

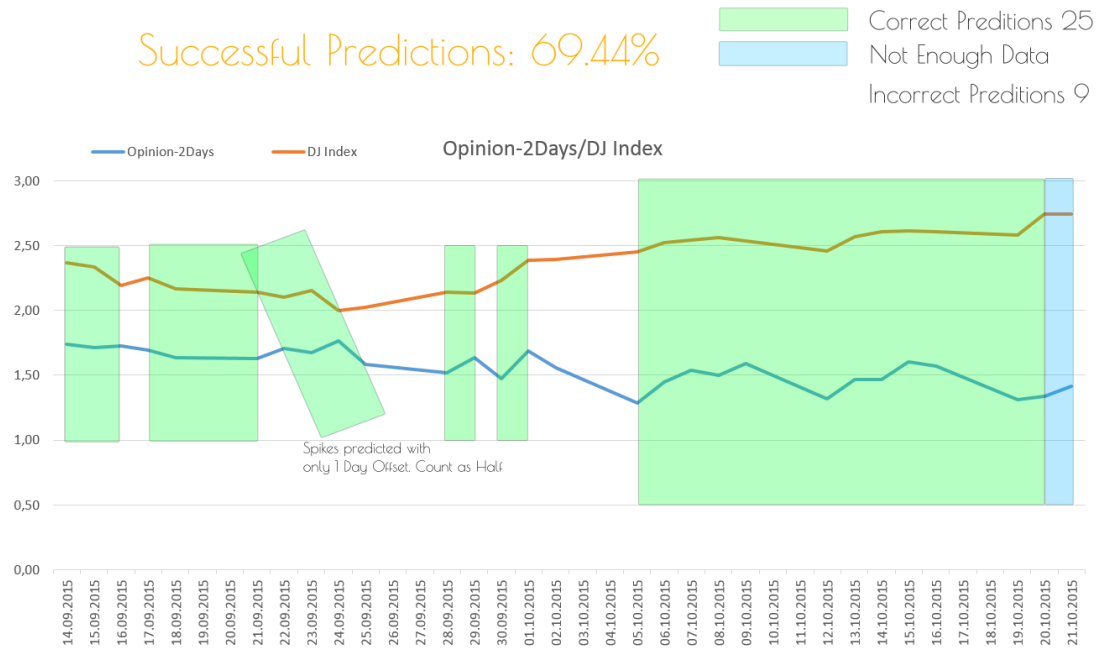

Image 2. Results of Dow Jones Prediction Two Days Ahead

Correct Predictions	Incorrect Predictions	Total Days	Prediction Accuracy
25	9	36	69.44%

NASDAQ Index

The smallest amount of success in predicting the Stock Market Indexes was for NASDAQ. this index as well had an almost 100% rate of prediction for the last week. Since no data was available for the 23rd of October, the last day of the graph was not taken into consideration.

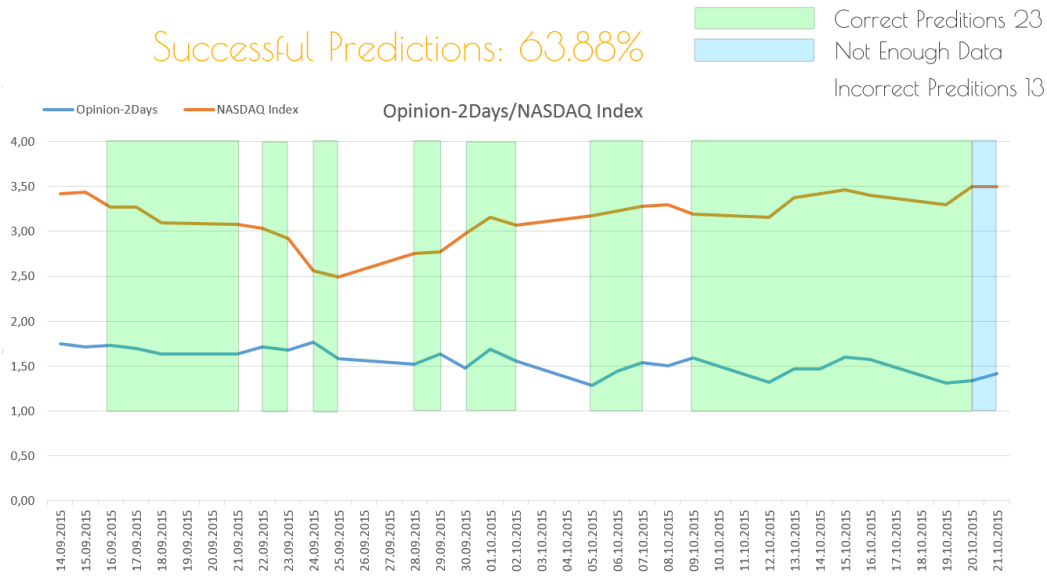

Image 3. Results of NASDAQ Prediction Two Days Ahead

Correct Predictions	Incorrect Predictions	Total Days	Prediction Accuracy
23	13	36	63.88%

S&P 500 Index

The S&P 500 index was predicted with an accuracy rate of 66.66%. That is 2 out of 3 predictions made by the algorithm were correct. As with the previous indexes, the last two weeks showed a very high accuracy, missing only two out of 16 days. The issue with this was that it could only predict the Stock Market fluctuation only a day ahead.

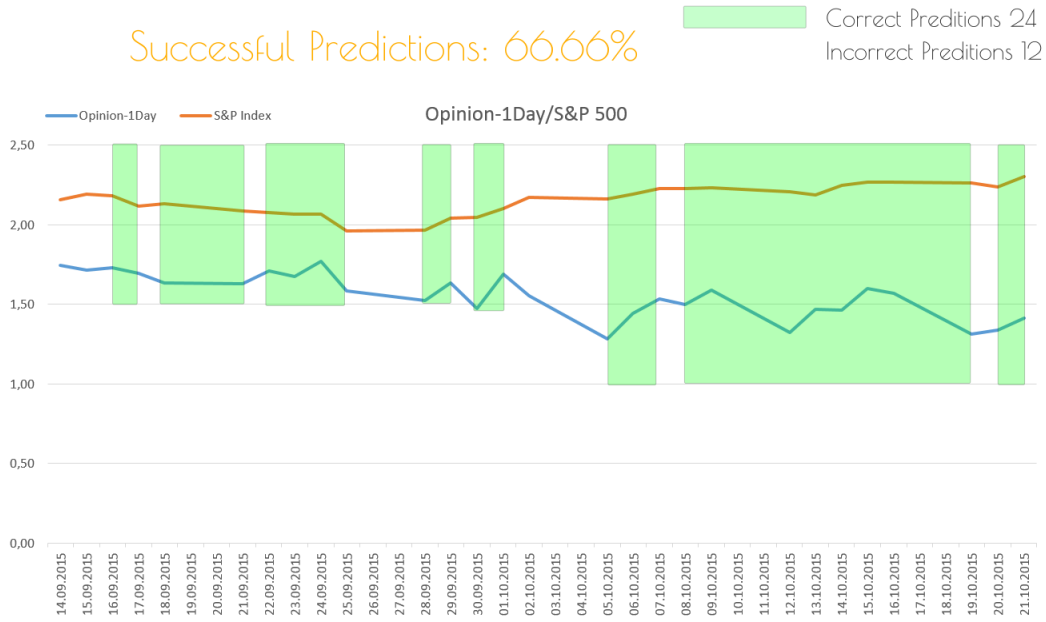

Image 4. Results of Dow Jones Prediction Two Days Ahead

Correct Predictions	Incorrect Predictions	Total Days	Prediction Inaccuracy
24	12	36	66.66%

5. Conclusions

As predicted, Stock Markets are heavily influenced by the news, and therefore, a sentiment analysis on them can indeed predict the fluctuation with a medium degree of accuracy. For our experiment, the medium prediction was the following

Total Correct Predictions	Total Incorrect Predictions	Prediction Accuracy	Prediction Inaccuracy
72	34	66.66%	31.48%

All indexes had an almost 100% rate of prediction for the last week. Since the algorithm is not yet in its final form, we can expect these numbers to go higher. It also helps the fact that

the more data goes into the system, the more the NewsInn Algorithm learns, and the better associations it can do.

6. Further Research

In order to thoroughly test the Artificial Intelligence Algorithm, and its prediction on the stock market, we will need to wait some more time, so that it can collect more news data. Even if it was tested against 24,000 articles, we do not feel this has enough significance. We will keep monitoring and hope that, in six months' time, NewsInn will have collected enough information to properly test it.

References

NewsInn Website http://newsinn.org/
SentiWordNet Website http://sentiwordnet.isti.cnr.it/
Wikipedia - Stock Market Index Page https://en.wikipedia.org/wiki/Stock_market_index
Yahoo Finance Historical Quites https://uk.finance.yahoo.com/q/hp?s=%5EDJI
S&P http://www.standardandpoors.com/en_US/web/guest/home?pagename=spcom
/template/topic _template/DescriptionPglt&r=US&l=EN&ds=indices_totalreturn

Copyright Disclaimer