Prediction of stock values changes using sentiment analysis of stock news headlines

László Nemes & Attila Kiss

Prediction of stock values changes using sentiment analysis of stock news headlines

László Nemes ⬤[a] and Attila Kiss ⬤[a,b]

[a]Department of Information Systems, ELTE Eötvös Loránd University, Budapest, Hungary; [b]Department of Informatics, J. Selye University, Komárno, Slovakia

ABSTRACT

The prediction and speculation about the values of the stock market especially the values of the worldwide companies are a really interesting and attractive topic. In this article, we cover the topic of the stock value changes and predictions of the stock values using fresh scraped economic news about the companies. We are focussing on the headlines of economic news. We use numerous different tools to the sentiment analysis of the headlines. We consider BERT as the baseline and compare the results with three other tools, VADER, TextBlob, and a Recurrent Neural Network, and compare the sentiment results to the stock changes of the same period. The BERT and RNN were much more accurate, these tools were able to determine the emotional values without neutral sections, in contrast to the other two tools. Comparing these results with the movement of stock market values in the same time periods, we can establish the moment of the change occurred in the stock values with sentiment analysis of economic news headlines. Also we discovered a significant difference between the different models in terms of the effect of emotional values on the change in the value of the stock market by the correlation matrices.

1. Introduction

A popular goal is to develop and/or use a model to sentiment prediction by looking for connections between words and marking them with positive or negative sentiments. There are many opportunities these days to perform sentiment analyses, for example external services that are almost completely ready to use it in a given context where it is needed like TextBlob. In addition, there are options that allow us to create our own models, train them based on our own data. Sentiment analysis with BERT is one of the most powerful tool that we can use, but we can also create a Recurrent Neural Network (RNN) as well or use the NLTK tool with VADER Lexicon with SentimentIntensityAnalyzer.

The stock market is one of the most important economic participants. Many people try to interpret and define the different stock market movements in many ways. In this article, we use different tools to the sentiment analysis, especially focussing on the economic news, but in terms of economic news, focussing only on the headlines of economic news. In today's communications and news consumption, the headlines of various articles play an even more important role than before. Now, we use sentiment analysis on these headlines on a particular company or companies to determine the effects of the headlines to the stock market. The question arises how much effect has the economic headline without the economic news whole context, if it has any measurable effect at all. We have found that it really has. Thus, we define the different impacts and their perceived significance with a very specific and unique new approach.

Data is an important pillar of analysis. Primarily the headlines of economic news are needed, what we use for sentiment analysis. Secondary, different stock market data are also needed based on companies. There are many possibilities for data collection and analysis, from 'traditional' dictionary-based performed by humans to 'more serious' neural networks that determine the polarity of the headlines of each economic news and label with appropriate emotional polarity. In the case of stock market data, numerous tools are available to obtain stock market data which can be even company-specific which is important to us. In both cases, we work with the most up-to-date data as possible, based on the information provided by the companies. Both, the headlines of the economic news and stock value data are related to the time period which specified by the news. Thus, the results of the given emotional analysis and the range of stock market data will be appropriate.

The analysis can be separated to the next sections. Collect headlines of economic news based on companies and collect stock market data according to the timestamps of the given economic news headlines.

Then prepare these data and apply different sentiment analysis tools like RNN or NLTK with VADER Lexicon ect. The RNN model was built and taught using the libraries and capabilities provided by Tensorflow. Manage these data and compare the stock market data and emotional data with visualization and explanation. Present how the headlines of economic news can affect different stock market changes and the public.

2. Related works

Devlin et al. (2018) introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers, that was designed to pretrain deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. The new possibilities and results of this model enable even low-resource tasks to benefit from deep unidirectional architectures. This model became one of the most significant tool of the natural language processing.

Wang et al. (2020) introduces a public sentiment analysis during the outbreak which is able to provides insightful information in making appropriate public health responses. They analyze the Sina Weibo popular Chinese social media site posts, where the unsupervised BERT model is adopted to classify sentiment categories (positive, neutral, and negative) and TF-IDF (term frequency-inverse document frequency) model is used to summarize the topics of posts. Analyzing posts with negative sentiment from social

media could contribute to understanding the experiences and offers examples for other countries. The analyses provide insights on the evolution of social sentiment over time and the topic themes connected to negative sentiment on the social media sites. BERT classification model and TF-IDF topic extraction model results were delivered with considerable accuracy.

The big data is a very popular and powerful tool nowadays. Lee (2020) explores the initial impact of COVID-19 sentiment on US stock market using big data notedly Daily News Sentiment Index (DNSI) and Google Trends data on coronavirus-related searches. The goal is to investigate a correlation between COVID-19 sentiment and 11 selected sector indices of the Unites States (US) stock market in a declared time period. Any positive or negative sentiment of public related to stock market crisis can have a ripple effect on decision making by investors in stock markets. The results reveal the distinct effects of the COVID-19 sentiment across in various industries and separate them to different correlation groups.

Khedr and Yaseen (2017) aims at constructing an effective model to predict stock market future trends with small error ratio and improve the accuracy of prediction. Where this prediction model is based on sentiment analysis and historical stock market prices, worked with K-NN and naïve Bayes algorithm to earn the final results. We can separate the model for two stages. The first stage is to determine the news polarity is positive or negative using naïve Bayes algorithm, the second stage incorporates the output of the first stage as input with the processed historical numeric data to predict the future stock trend using K-NN algorithm.

Streaming data prove to be a rich source of data analysis where data are collected in real-time. The major characteristics of such data being its accessibility and availability, help in proper analysis and prediction. Das et al. (2018) show an analysis that has been made for making financial decisions such as stock market prediction, to predict the potential prices of a company's stock using twitter data.

Kalyani et al. (2016)'s project takes data such as financial news articles about a company and predict its future stock trend with news sentiment classification, assuming that news articles have impact on stock market. This is an attempt to study relationship between news and stock trend. For this, they used dictionary based approach. The dictionaries for positive and negative words are created using general and finance specific sentiment carrying words. Based on this data, they implemented classification models. The results show that Random Forest (RF) and Suport Vector Machine (SVM) perform well in all testing.

Mikolov et al. (2011) presents some modifications of the original recurrent neural network language model (RNN LM). This model has been shown to significantly outperform many competitive language modelling techniques in terms of accuracy, but the remaining problem is the computational complexity. Their result is more than 15 times faster in both training and testing phases. The resulting RNN model can be smaller, faster and more accurate than the base.

In another paper we can get to know the SummaRuNNer model, which is a Recurrent Neural Network (RNN) based sequence model. Nallapati et al. (2016) proposes a very interpretable neural sequence model for extractive document summarization that allows intuitive visualization, and shows that it is better performing than the state-of-the art deep learning models and it is comparable to this learning models as well.

Following this line, Liu et al. (2016) shows the multitask learning framework to jointly learn across multiple related tasks which based on recurrent neural network. They

propose three different mechanisms of sharing information to model text with task-specific and shared layers where the differences among them are the mechanisms of sharing information between the tasks.

Let's look at other approaches. Balahur (2013) presents a method for sentiment analysis specifically designed to work with Twitter data, focussing their structure, length and specific language. They show that the use of generalized features significantly improves the results of the sentiment classification. They apply unigram and bigram (n-gram) and supervised learning with simple Support Vector Machines. Based on the results we can conclude that, the best properties to use emotional analysis is the unigram and the bigram together. We can also see that generalizations, using unique tags, emotive words and modifiers are strongly improve the performance rating of emotions.

SmartSA is a lexicon-based sentiment classification system for social media genres by Muhammad et al. (2016). It integrates strategies to capture contextual polarity from two ways, the interaction of terms with their textual neighbourhood and text genre like local and global context. They also introduce an approach to hybridise a general purpose lexicon, with genre-specific vocabulary and sentiment. The results from diverse social media show that this strategies of local and global contexts significantly improve sentiment classification, and are complementary in combination.

Arras et al. (2017) have introduced a simple yet effective strategy for extending the LRP procedure to recurrent architectures (LSTM) by proposing a rule to backpropagate the relevance through multiplicative interactions. They applied the extended LRP version to a bidirectional LSTM model for the sentiment prediction of sentences.

To study the influence of market characteristics on stock prices, traditional neural network algorithms may incorrectly predict the stock market, since the initial weight of the random selection problem can be easily prone to incorrect predictions. Based on the development of word vector in deep learning, Pang et al. (2020) demonstrates the concept of 'stock vector.' The input is not only a single index or single stock index, but multi-stock high-dimensional historical data. They propose the deep long short-term memory neural network (LSTM) with embedded layer and the long short-term memory neural network with automatic encoder to predict the stock market.

Billah et al. (2016) presented an improved Levenberg Marquardt (LM) training algorithm. Improved Levenberg Marquardt algorithm of neural network can predict the possible day-end closing stock price with less memory and time needed, provided previous historical stock market data of Dhaka Stock Exchange. such as the opening, highest, lowest prices and total share traded data.

3. DataFrame building

3.1. Options to build DataFrame of the news headlines and stock values

There are several ways to approach data structure building. Primarily we consider the headlines of economic news. Of course, there is the possibility to compile a collection of data by human effort according to specific conditions, such as gathering economic news titles filtered by a given company name from the collection built from a start time (which is the oldest possible economic news titles) until to reach a certain limit. There is the possibility of approaching the analysis using data from a previous archive collection of data, but the main goal is to use the most up-to-date data as possible. There is

also the possibility of using human effort in the case of data collection from the stock market values of companies, but today many economic portals and other libraries and frameworks are available to fully automate the process. In this case, automation plays a more important role than in the previous economic news title data collecting. The error factor can be significantly reduced when compiling companies' economic data. In addition, the source and the values of the stock data are easier to manage this way than in the economic news title data collecting.

3.2. DataFrame of the headlines of economic news

As mentioned earlier, the main goal in the headlines of economic news is to use the most up-to-date data possible. All data collection and management is automated. There is an option to the user to specify the portal as a source to manage the news. We used data from 'finviz.com' for our analyses. Before collecting the data, it is possible to enter the stock exchange names of the companies where we would like to collect the data of recent economic news for analysis. It is possible to specify more than one company by listing as parameter. The function takes care of managing the appropriate timestamps (news publication time) and separating the news based on the companies and create a backup into a file as csv. This freshly compiled data is used by the application for further analysis (as part of sentiment analysis, comparisons, and other possibilities.) It is important to mention that news timestamps play a role in compiling additional stock market data so the analyses take place in the same time period. Thus, these economic news headlines define the interval for subsequent stock market data collection separated for companies.

```
website_url = 'https://finviz.com/quote.ashx?t='
company_tikcers =[ 'AMD' , 'AMZN' , 'FB' , 'GOOG']
news_tables ={}
parsed_data =[]
for ticker in company_tikcers:
  url = website_url + ticker
  req = Request(url=url, headers={ 'user-agent' : 'my-scrape'})
  response = urlopen(req)
  html = BeautifulSoup(response, 'html')
  news_data = html.find(id='news-table')
  news_tables[ ticker] = news_data
for ticker, news_table in news_tables.items():
  for row in news_table.findAll('tr'):
    title = row.a.text
    date_data = row.td.text.split('')
    if len(date_data) == 1:
      time = date_data[ 0][ 0:7]
    else:
      date = datetime.datetime.strptime(date_data[ 0], '%b-%d-%y').strftime('%Y/%m/%d')
      time = date_data[ 1][ 0:7]
```

```
    parsed_data.append([ ticker, date, time, title])
  dataset  =  pd.DataFrame(parsed_data,  columns=[ "Company",
"Date", "Time", "News Headline"])
```

Listing 1. Part of the Economic news headlines dataframe builder

The code snippet shown by Listing 1 implements a part of data collection for economic news headlines. Where the 'weblite url' is the portal from where we process the news, and the 'company tickets' are the company names on the stock market in a list from which we would like to compile data. The data processing shown by the code snippet use the 'BeautifulSoup', 'urlopen' and 'Request' tools to perform scraping. For other source pages we have to make changes in this processing stage to scrape data from this specified page (Figure 1).

3.3. DataFrame of the company specific stock values

'Yahoo fin' tool was used to collecting stock values for companies. This data is separated by companies into the intervals of previously compiled economic news headlines. Based on this, it will be possible to analyze and compare economic news headlines and stock market data for a given period. This data collection and management also provides the opportunity to perform individual and aggregate analyses (Figure 2).

4. Sentiment analysis with different tools

As mentioned earlier, there are many possibilities for sentiment analysis from human-labelled data to various deep learning methods. In the present case, we compare the possibilities offered by TextBlob, NLTK -- VADER Lexicon, RNN and BERT. The main goal is to analyze the headlines of economic news about different companies and determine their sentiment values to be positive or negative possibly neutral. A key factor is to minimize neutral values. It should be noted that we do not have as much influence over external third-party devices as we do about our own models, such as RNN.

In the case of sentiment analysis, the headline of the economic news from each company is labelled to what sentiment value it carries, and the polarity value is also indicated. With the help of these data, we can make a number of further analyses and

Figure 1. Part from the economic news headlines dataframe.

open	high	low	close	adjclose	volume	ticker	date
82.55000305175781	84.97000122070312	80.86000061035156	82.2300033569336	82.2300033569336	69423700	AMD	2020.10.26
82.0	82.37000274658203	77.56999969482422	78.87999725341797	78.87999725341797	156669500	AMD	2020.10.27
78.7300033569336	78.95999908447266	75.76000213623047	76.4000015258789	76.4000015258789	76529900	AMD	2020.10.28
76.75	79.18000030517578	76.29000091552734	78.0199966430664	78.0199966430664	52784100	AMD	2020.10.29
77.08999633789062	77.69999694824219	74.2300033569336	75.29000091552734	75.29000091552734	51349000	AMD	2020.10.30
75.8499984741211	76.33999633789062	73.76000213623047	74.69999694824219	74.69999694824219	45760700	AMD	2020.11.02
74.93000030517578	77.08000183105469	74.5999984741211	76.58000183105469	76.58000183105469	41141400	AMD	2020.11.03

Figure 2. Part from the AMD Stock Values DataFrame.

comparisons. The main direction is to compare the specific companies with their stock market values in the period of time which determined by the economic news. Thus assessing and presenting the emotional impact of economic news headlines on stock market changes and see how powerful the headlines can be alone without full content.

In addition, our goal is to determine the strength and accuracy of different sentiment analysis tools by the given context. The BERT tool is used as a kind of comparative tool to see how close the results of the other tools are to the results of BERT. More detailed analysis of stock market values and sentiment values (polarity and sentiment label) is done using the results of TextBlob, NLTK -- VADER Lexicon and RNN.

4.1. TextBlob

TextBlob is a powerful NLP library for Python, which is built upon NLTK and provides an easy to use interface to the NLTK library. This tool can be used to perform a variety of NLP tasks ranging from parts-of-speech tagging to sentiment analysis, and language translation to text classification, but we focus on the sentiment analysis. If we do a sentiment analysis, we actually determine a polarity value of the sentences, where this value can be between −1 and 1. Then we label the data with the right sentiment value (positive, negative or neutral). For other tools, the polarity value may move on a different scale, so the labelling needs to adjust for these differences for further analysis.

The Figure 3 shows that sentiment values separated by companies. No other value can approach the neutral section, it can be concluded that the analysis of the given economic news headlines and its outcome is very uncertain. In the case of AMD, it can be noted that in Figure 3(a), in addition to the 63 news headlines rated as neutral, 31 are positive and 5 are negative. In the case of FB -- Facebook, in addition to the 80 news headlines rated as neutral, there are 13 positive and 6 negative values as well.

In the case of the total result, 75.25 percent in Figure 3(b) is neutral besides to this 20.25 percent is positive and only 4.50 percent is negative.

The goal is to minimize neutral values by using a more accurate analysis to reduce the inaccuracy increased by its neutral values in stock market comparisons.

The following figure (Figure 4) shows the results divided into days in the interval. The results are aggregated and this gives us a normalized value of how positive or negative the overall day was for the company. Due to the significant neutral value of more than 75 percent, the days are visibly shifting in a positive direction, which can greatly distort real results. Where a company does not have a coloured column for a given day, there was no economic news headline about those company. The following figure is formed on the interval, where above zero means the positive section and below means the negative section.
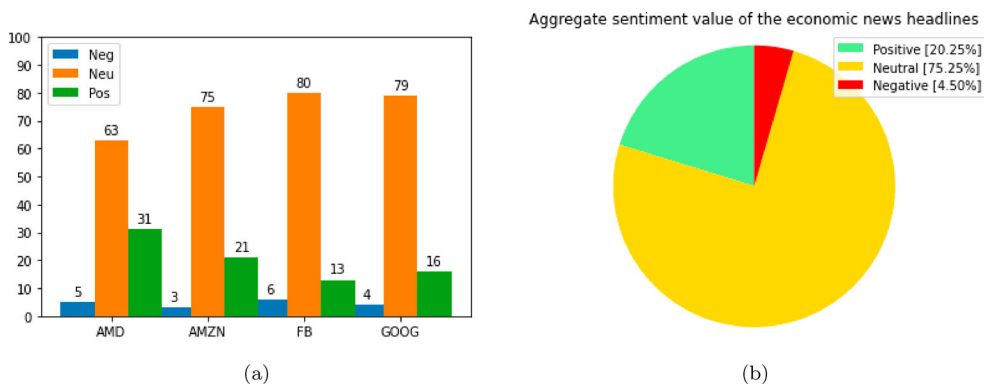
Figure 3. Company specific results of the sentiment analysis using TextBlob. The time period stands between 2020-10-27 and 2020-11-14. (a) Results by Companies and (b) Aggregate Sentiment Result.

It should be noted that there were a large amount of news during the period, about AMD will launch new CPUs and GPUs in October, which was also significantly positive. This may be explained by recent period in the end of October CPU and GPU events and this is the effect of these events.

4.2. NLTK -- VADER lexicon

NLTK stands for Natural Language Toolkit. This toolkit is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to it with an appropriate response. Again, we focus on sentiment analysis with the SentimentIntensityAnalyzer. The polarity value of the sentences scales between -1 and 1 just like in the TextBlob. The data labelling process (positive, negative or neutral) is similar to the previous tool. We use VADER Lexicon in this section. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis
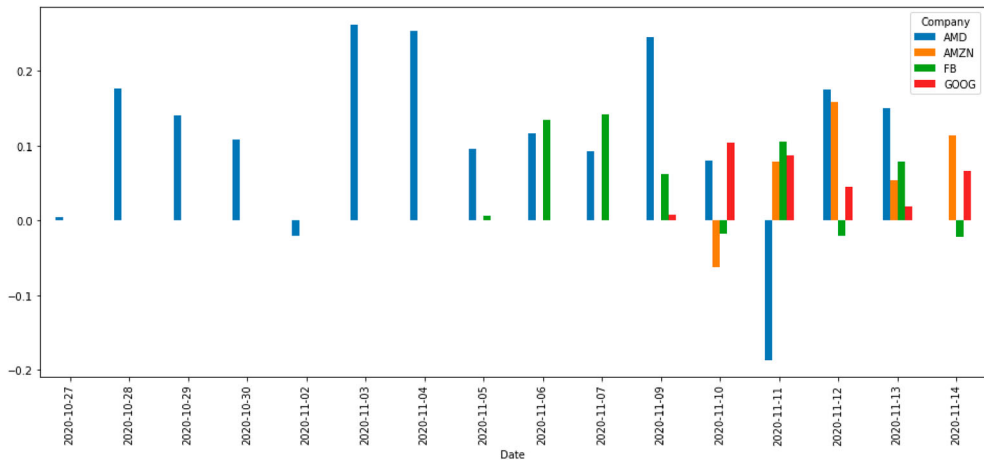

Figure 4. TextBlob Analysis results separated by days. The time period stands between 2020-10-27 and 2020-11-14.
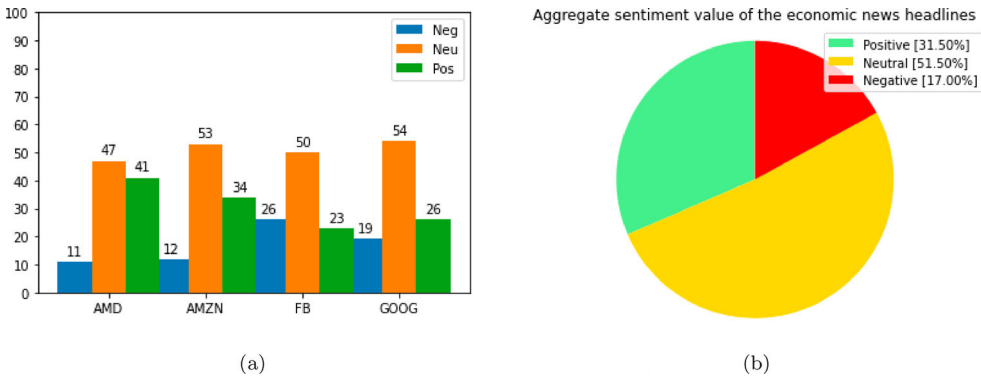
Figure 5. Company specific results of the sentiment analysis using NLTK -- VADER Lexicon. The time period stands between 2020-10-27 and 2020-11-14. (a) Results by Companies and (b) Aggregate Sentiment Result.

tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains.

As shown in Figure 5 below, the neutral value (a) dominates in all cases among the sentiment results separated by companies. In the figure next to it (b), the aggregate sentiment result shows the economic news headlines significant neutral values. This level of neutral values has impact on comparisons and analyses to the subsequent stock market changes. Compared to the results of TextBlob, the neutral values have been significantly reduced and we expect that has significant effect in further analysis to obtain more accurate and realistic results with fewer neutral values. In Figure 5(b), 51.50 percent of the total result is neutral in addition to 31.50 percent positive and 17 percent negative. Of the positive or negative categories, the positive strongly dominates, but this huge neutral value still makes the result little bit uncertain.

The following figure (Figure 6) shows the results divided into days in the interval. The results are aggregated and this gives us a normalized value of how positive or negative the overall day was for the company. One day in total cannot be neutral because of the other news headlines have to move it in some direction and the neutral values according to the polarity also try to move the result in some direction too. Thus, the following figure is formed on the interval, where above zero means the positive section and below means the negative section.

4.3. Recurrent neural network (RNN)

When we talk about traditional neural networks, all the outputs and inputs are independent of each other. But in the case of recurrent neural networks, the output from the previous steps is fed into the input of the current state.

All in all the Recurrent Neural Network is a neural network that is intentionally run multiple times, where parts of each run feed into the next run. Specifically, hidden layers from the previous run provide part of the input to the same hidden layer in the next run. Recurrent neural networks are particularly useful for evaluating sequences, so that the hidden layers can learn from previous runs of the neural network on earlier parts of the sequence.
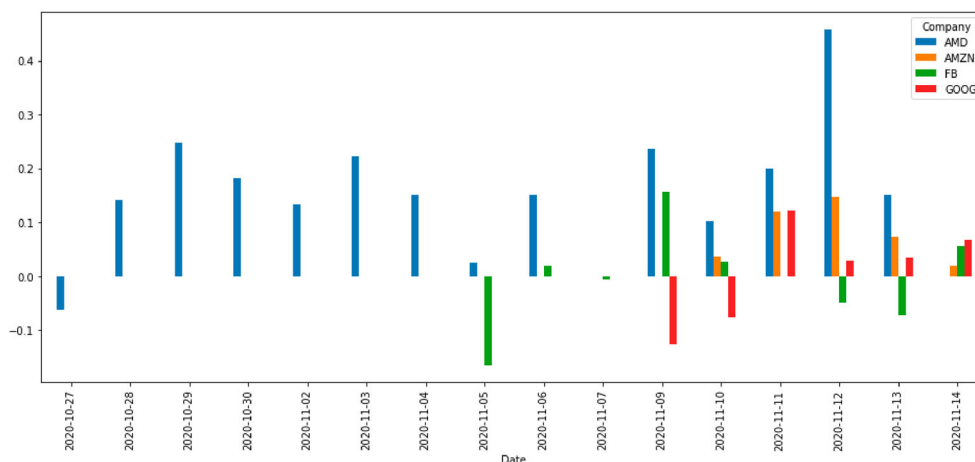
Figure 6. NLTK -- VADER Lexicon Analysis results separated by days. The time period stands between 2020-10-27 and 2020-11-14.

For example, the following figure of Google shows a recurrent neural network that runs four times (Figure 7). Notice that the values learned in the hidden layers from the first run become part of the input to the same hidden layers in the second run. Similarly, the values learned in the hidden layer on the second run become part of the input to the same hidden layer in the third run. In this way, the recurrent neural network gradually trains and predicts the meaning of the entire sequence rather than just the meaning of individual words.

An advantages of the RNN model: RNN can process inputs of any length. An RNN model is modelled to remember each information throughout the time which is very helpful in any time series predictor. Even if the input size is larger, the model size does not increase. But there some disadvantages: Due to its recurrent nature, the computation is slow. Training of RNN models can be difficult.

We have to mention that the polarity value of the sentences scales between 0 and 1 here. In contrast to the models what mentioned earlier.

The Figure 8 shows the results of RNN separated by companies and the aggregating result as before. A significant difference from the previous results of TextBlob and NLTK -- Vader Lexicon is that the neutral section was completely eliminated, all news headlines were categorized as either positive or negative. This is a significant difference from previous models, although there was a kind of downward trend in the models. The neutral category of the TextBlob was huge, it was significantly reduced by the NLTK -- Vader Lexicon, and then the RNN model was managed to avoid a neutral category.

Figure 8(a) shows how the positive and negative news headlines are distributed among the companies. In the case of AMD, it can be noted that the result is quite 'balanced' with 51 positive and 48 negative values. In part (b) of the figure, the total result is 58.50 percent positive and 41.50 percent negative and the neutral value is 0 percent which is now the key.

In Figure 9, the positive negative day categorization is totally different than the previous ones, because in the case of the RNN model, the polarity values scale between 0 and 1. Therefore, here is a 'traditional' bar chart showing the aggregation of polarity values for each day.
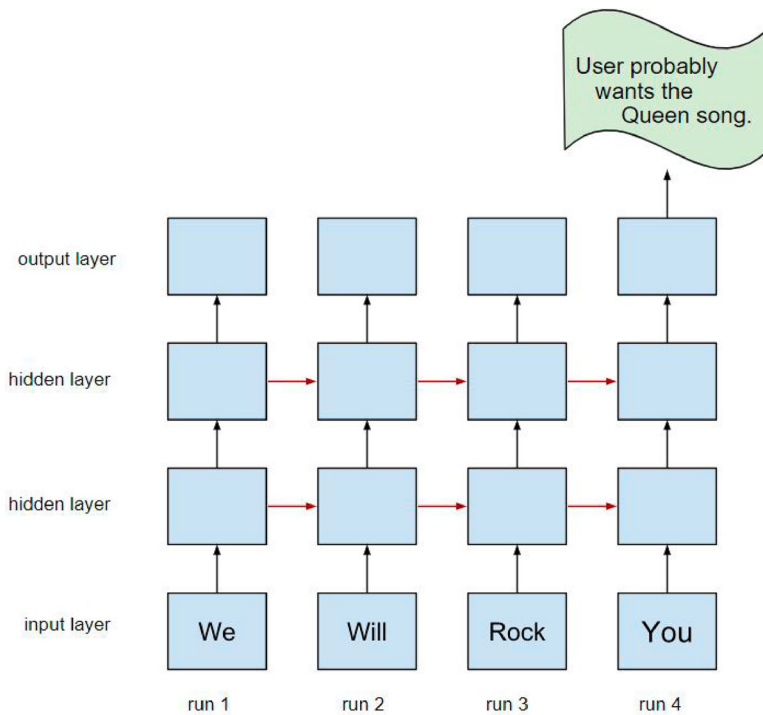
Figure 7. A recurrent neural network.[1]

Note: The RNN model was trained based on an IMDB review dataset[2] (In the test and train dataset sections we used shuffle method as well. Then we use the fresh scraped dataset as test dataset with this trained model.)

4.4. Bidirectional encoder representations from transformers (BERT)

Unlike the traditional NLP models that follow a unidirectional approach, that is, reading the text either from left to right or right to left, BERT reads the entire sequence of
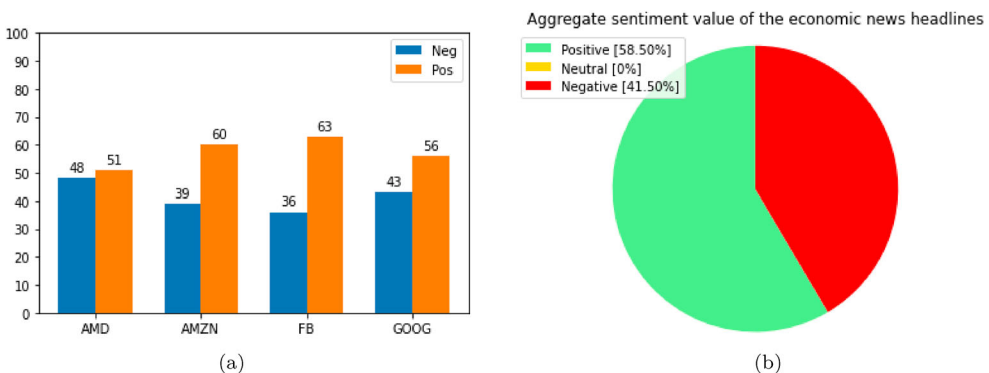

Figure 8. Company specific results of the sentiment analysis using RNN. The time period stands between 2020-10-27 and 2020-11-14. (a) Results by Companies and (b) Aggregate Sentiment Result.
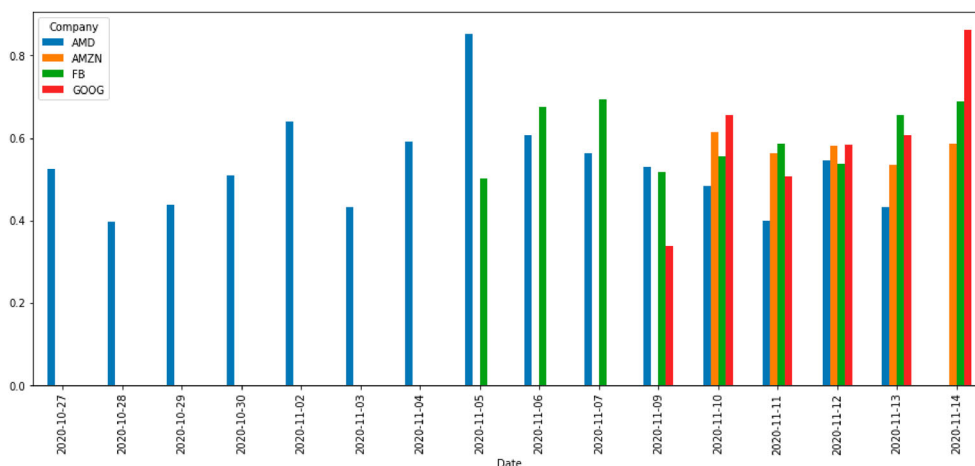
Figure 9. RNN Analysis results separated by days. The time period stands between 2020-10-27 and 2020-11-14.

words at once. BERT makes use of a Transformer which is essentially a mechanism to build relationships between the words in the dataset. In its simplest form, a BERT consists of two processing models -- an encoder and a decoder. The encoder reads the input text and the decoder produces the predictions. But, because the main goal of BERT is to create pre-trained model, the encoder takes priority over decoder. BERT is a remarkable break-through in the field of NLP.

As mentioned earlier, BERT is used as a kind of comparative result. Figure 10 shows the results obtained by BERT. Of course, without a neutral category, it managed to categorize each economic news headline and labelled it as a positive or negative value. In part (a) of the figure, it can be mentioned that the result of our previous RNN model is quite encouraging, as there is no neutral category either and the values of certain companies are quite close to the result of BERT. Part (b) of the figure shows the overall result where 50.50 percent is positive and 49.50 percent is negative compared to the result of the RNN model where 58.50 is positive and 41.50 is negative, neutral is 0 percent in both cases.

We expected that the model we trained and taught would give more accurate and more reliable results than other tools on the same data set. More specifically, the result from the RNN model determines emotional values and labels with a more accurate and smaller error rate than NLTK with VADER Lexicon or TextBlob. This expectation was also confirmed by the results. It should be emphasized that the result of NLTK was much more encouraging than initially expected and in later analyses, despite the existing neutral values, it gave a much 'finer' result than TextBlob where we get a 'raw' result due to the significant neutral value. For the RNN model, no headline is placed in the neutral category. Regarding the results of BERT and the results of the other tools, we expect more accurate results from the RNN and NLTK tools when analyzing with stock market values.

5. Sentiment and stock value analysis

Following the sentiment analyses at a given interval, we can start the comparison with stock market changes at the same interval. During the sentiment analysis, the 'realistic'
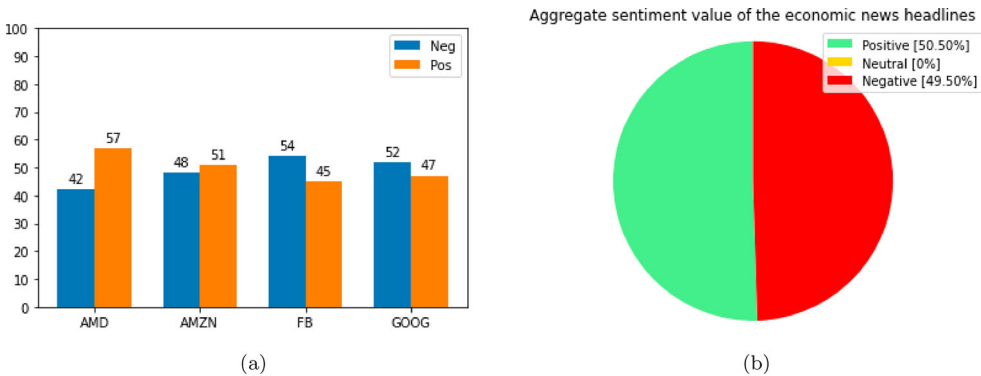
Figure 10. Company specific results of the sentiment analysis using BERT. The time period stands between 2020-10-27 and 2020-11-14. (a) Results by Companies and (b) Aggregate Sentiment Result.

word was mentioned, which refers to the smaller neutral category, less neutral value in the solution of the analysis. It refers to a better and 'stronger' analytical model that was able to give positive and negative tags to news headlines which were categorized as neutral by the previous analytical model. Thus we reduced the potential of error and possible skew results.

We can say that economic news do have an impact on stock market shifts, there are times when certain news items have effect to the later movements, and there are times when the news describe a particular shift, which enhances change too.

Our main study in the present case focuses on the headlines of economic news about the various companies which was given as parameters previously, without their full article context. The headline itself, which aims to draw people's attention and generate clicks on full content, is worded in this 'sometimes sharp, eye-catching' way. How much impact do these economic news headlines have on stock market changes, if it has any effect. In our results we found that it really has.

Figure 11 shows the AMD stock market changes during the given study period, where the date and the daily closing (adjusted closing price) value are displayed.

The following (Figure 12) shows the results of different sentiment models (TextBlob, NLTK -- Vader Lexicon and RNN) for the given period, broken down by day.

Here, the results are the same of the previous sentiment analyses, but now they are displayed on a different diagram for the purpose of being comparable with the stock market data. Significant differences can be observed in the results of the different models especially in some parts of the result. One of the most striking may be the negative news stream around 2020-11-11. In all three cases, a negative trend can be detected, but the differences in the extent are significant. These results, in comparison with stock market changes, help us to see a kind of effect on whether stock market movements are reflected in the diagram of sentiment results. The amount of neutral values plays a significant role in the accuracy of the models. It was mentioned earlier that when calculating daily results (this day is positive or negative all in all), the polarity values of the neutral values also count, so that these values also play a role in the positive or negative shift of a day as they belong to that day, but this values distort the result. In contrast, in a model where there is no neutral value, much higher accuracy can be expected.
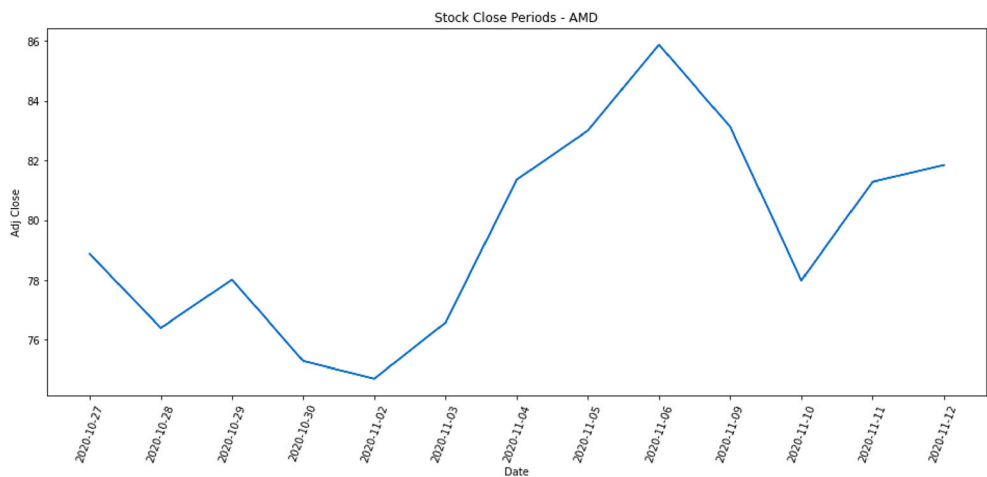
Figure 11. AMD stock value changes.

Figure 13 shows the normalized results, where the different models show how the stock market value changed in the period and how the daily results obtained by the economic news headlines of the given period relate to stock market movements. The graph still process data from AMD. The first figure (a) shows the summary of the results obtained by the TextBlob and the stock market result. In the case of TextBlob, the ratio of neutral values was 75.25 percent, which is also reflected in the large 'vibration' of emotional results. On the normalization graph in this case with the economic news headlines and stock values we can read as fundamental changes, with the trend of decreasing or increasing. In the period between 2020-11-04 and 2020-11-06, a strong decrease in emotional values can be observed, in addition to with a smaller 'break' or a correction in the stock market values as well. In the phases of 2020-11-09 and 2020-11-10, a significant
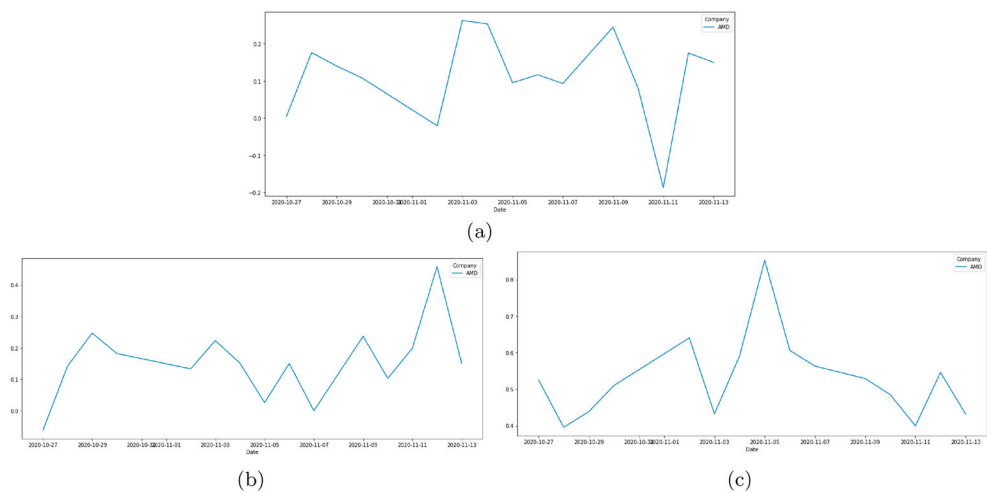

Figure 12. Sentiment analysis of different models by daily separation. (a) TextBlob. (b) NLTK -Vader Lexicon and (c) RNN.

break point can be observed in both stock market developments and emotional results. Overall, we can see the impact and the major growth declines can be traced from the chart, but its detail is questionable. In the case of figure (b) we can see the results of the NLTK -- Vader lexicon normalization. The ratio of neutral values in this case was reduced to 51.50 percent. It can be said that the result is surprising at first. It is clear that a more detailed 'co-movement' of stock and emotional values is shown in the figure. Changes between 2020-11-09 and 2020-11-11 will be tracked 'fully in sync'. Regarding the results of RNN in figure (c), where the ratio of neutral values was 0 percent, significant differences can be observed compared to the previous ones. Here, it may appear primarily that the two results do not follow each other in 'synchrony' and in some cases there is a significant difference between emotional and stock market results. It can be assumed that in this case, the effect of emotional values on the results of the current days may not be as great and perhaps a kind of 'periodic pre-diction' can be observed. The significant positive result between 2020-11-04 and 2020-11-06 is one of the most striking results. Until the subsequent correlation matrix results, all that can be stated is that there is a significant decrease in the influence of emotional values in the given stock market period. A kind of emotional decrease or increase and a following stock shift can be observed, but in fact the influence has decreased significantly, which can be explained by neutral values and 'realism,' when examining the influence of news headlines we cannot expect as much impact as full economic articles and analyses. In all three cases, these effects are also analyzed by correlation matrices.

In the case of Figure 14, we can see that Compound (sentiment results) has a huge impact on both the opening, closing, lowest and highest values of the stock market, which is a very distorted result. It's almost unthinkable to have such a big impact. As mentioned earlier, the significant neutral value can be traced back to this situation as well.
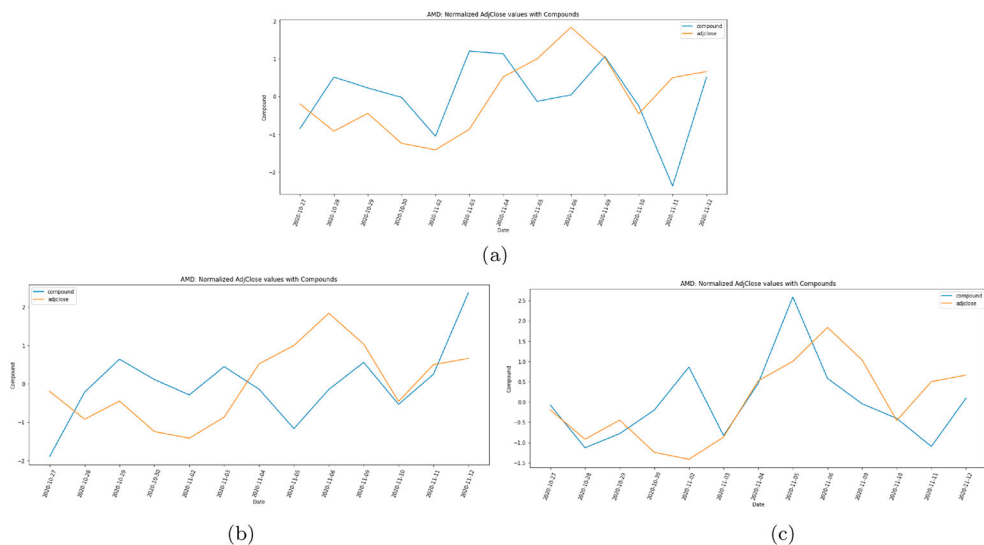

Figure 13. Normalized results of the sentiment and stock values. (a) TextBlob. (b) NLTK -Vader Lexicon and (c) RNN.

The following Figure 15 shows the result of the NLTK -- Vader Lexicon correlation matrix, where there are decreases in Compound values in almost all values compared to the results of the previous (TextBlob) correlation matrix. In addition to a kind of 'synchronized result' seen on previous diagrams, a significant effect was 'expected' in the matrix as well, but perhaps these results may also seem excessive as a result obtained, considering that we examine economic news headlines on a company-specific basis.

In the Figure 16, the correlation matrix of RNN is completely different and surprising in this case as well. The value of the Compound has decreased significantly compared to its previous models to the opening, closing, lowest and highest values, and unlike before, its effect on another value has increased drastically. The value of the volume is the amount of an asset or security that changes hands over some period of time, often over the course of a day.

The correlation matrix of RNN and the values of Compound provide a kind of explanation for the diagrams seen earlier. The previous models had a significant effect on the opening, closing, lowest and highest values, in contrast, the RNN shows a completely different result. Overall, we can say that the headlines themselves have a significant effect on the change in stock market values, in addition to highlighting the volume value, which alone received a significant value in the RNN model, unexpectedly high. It should be noted that the data from the study period may also play a role in this. But the result is thought-provoking. The result is not unique. We obtained a similar result for the measurements between 2020-10-27 and 2020-11-16 for another company, which was the Google (GOOG). As we can see in the Figure 17.

Figure 14. Correlation matrix of TextBlob.

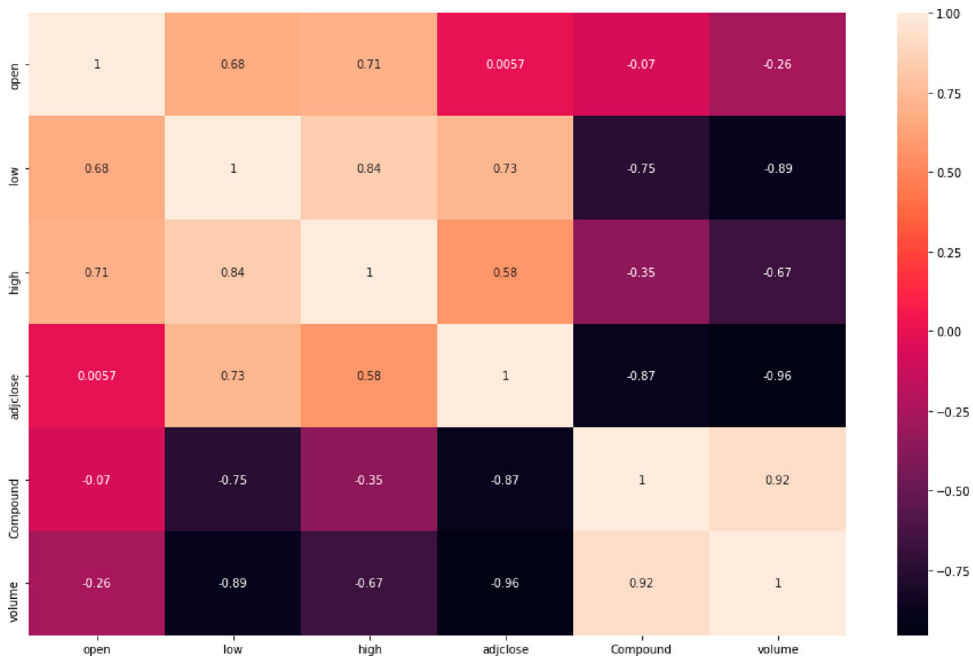**Figure 15.** Correlation matrix of NLTK -- Vader Lexicon.

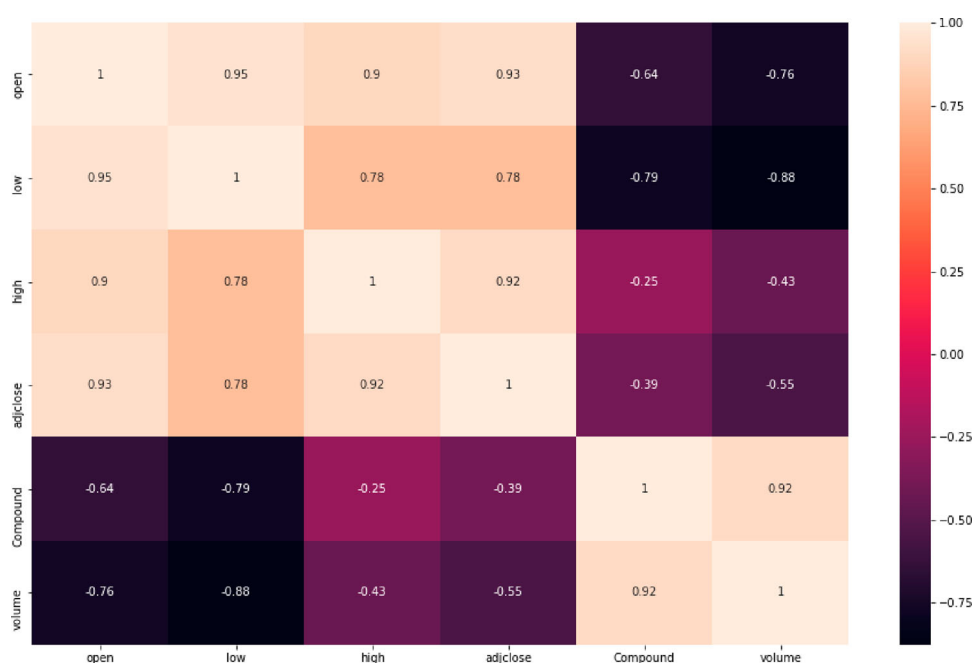**Figure 16.** Correlation matrix of RNN.

Figure 17. Correlation matrix of RNN with another company (GOOG).

6. Conclusion and future work

In this work, we used different sentiment analysis tools to emotionally analyze and classify different economic news headlines and examine their impact on different stock market value changes even without their full context. Emotions were classified into the usual positive negative and neutral categories. Neutral categories appeared for TextBlob and NLTK-VADER Lexicon tools, but not for Recurrent Neural Network (RNN). The various sentiment analyses results were compared with the result of BERT as a benchmark. As we expected, the results of the RNN model what we developed and taught outperformed the other sentiment analysis tools and gave a result quite close to BERT, emphasizing that there was no neutral emotional value in this case either. In the analysis of emotional results and stock market changes, we compared the daily results of emotional values and the results of stock market values for the given period. We obtained appropriate diagrams for the reading of the emotional results and the stock market movements and corrections, but we could detect differences according to the ratio and effect of the neutral values of the different models. In the field of further analysis, we detected significant differences in the correlation matrices. In the case of TextBlob, the Compound (emotional results) had a significant effect on the opening, closing, highest and lowest values of the stock exchange, the NLTK -- Vader Lexicon gave similar results, but reducing the results of the previous model significantly. The RNN model brought a completely different value. The emotional values and stock market change diagram also showed a kind of smaller effect, which was also confirmed by the correlation matrix, and also had a significant effect on the Volume value compared to the other models. Overall, economic news headlines have an impact on stock market values even without their textual context, and

significant differences can be observed between different sentiment analytical tools. But the stock market impact also depends on how the data in the current study period was affected.

Future work could include further expansion of the analyses, possible additions of a new features. In addition, the inclusion of other tools to compare stock market predictions with different sentiment analysis tools. That can be built into an easy-to-use format by developing a platform incorporating various future changes of tensorflow into the current model.

Notes

1. https://developers.google.com/machinelearning/glossary/#recurrent_neural_network
2. https://www.tensorflow.org/datasets/catalog/imdb_reviews

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Notes on contributors

László Nemes received the B.Sc. degree in computer science from Eötvös Loránd University in 2020 and currently pursuing a M.Sc. degree. He is a Demonstrator with the Department of Media and Educational Technology, Eötvös Loránd University.

Attila Kiss was born in 1960. In 1985 he graduated (MSc) as mathematician at Eötvös Loránd University, in Budapest, Hungary. He defended his PhD in the field of database theory in 1991. Since 2010 he is working as the head of Information Systems Department at Eötvös Loránd University. Since 2015 he has been teaching at J. Selye University, Komárno, Slovakia, too. His scientific research is focussing on database theory and practice, security, semantic web, big data, data mining, artificial intelligence and bioinformatics. He was the supervisor of 7 PhD students. He has more than 160 scientific publications.

ORCID

László Nemes http://orcid.org/0000-0001-6167-9369
Attila Kiss http://orcid.org/0000-0001-8174-6194

References

Arras, L., Montavon, G., Müller, K. R.., & Samek, W (2017). Explaining recurrent neural network predictions in sentiment analysis. arXiv preprint arXiv:1706.07206.

Balahur, A. (2013, 14 June). *Sentiment analysis in social media texts*. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, GA (pp. 120–128). Association for Computational Linguistics.

Billah, M., Waheed, S., & Hanifa, A (2016, December 8–10). *Stock market prediction using an improved training algorithm of neural network*. 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh (pp. 1–4). IEEE. https://doi.org/10.1109/ICECTE.2016.7879611

Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia Computer Science*, *132*, 956–964. https://doi.org/10.1016/j.procs.2018.05.111

Devlin, J., Chang, M. W.., Lee, K., & Toutanova, K (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Kalyani, J., Bharathi, P., & Jyothi, P. (2016). Stock trend prediction using news sentiment analysis. arXiv preprint arXiv:1607.01958.

Khedr, A. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, *9*(7), 22. https://doi.org/10.5815/ijisa

Lee, H. S. (2020). Exploring the initial impact of COVID-19 sentiment on US stock market using big data. *Sustainability*, *12*(16), 6648. https://doi.org/10.3390/su12166648

Liu, P., Qiu, X., & Huang, X (2016). Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101.

Mikolov, T., Kombrink, S., Burget, L., Černocký, J., & Khudanpur, S (2011, May 22–27). *Extensions of recurrent neural network language model*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic (pp. 5528–5531). IEEE. https://doi.org/10.1109/ICASSP.2011.5947611

Muhammad, A., Wiratunga, N., & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-based Systems*, *108*, 92–101. https://doi.org/10.1016/j.knosys.2016.05.032

Nallapati, R., Zhai, F., & Zhou, B (2016). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. arXiv preprint arXiv:1611.04230.

Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020). An innovative neural network approach for stock market prediction. *The Journal of Supercomputing*, *76*(3), 2098–2118. https://doi.org/10.1007/s11227-017-2228-y

Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via bert model. *IEEE Access*, *8*, 138162–138169. https://doi.org/10.1109/Access.6287639