Article

Sentiment Analysis of Current Reports Texts with Use of Cumulative Abnormal Return and Deep Neural Network

Maciej Wujec [1,]

[1] AI Lab, Science and Technology Park in Opole, Technologiczna 2, 45-839 Opole, Poland; maciej.wujec@gmail.com

Abstract: The deep neural network - BERT model (Bidirectional Encoder Representations from Transformers) and the stocks cumulative abnormal return is used in this article to analyze the sentiment of financial texts. The proposed approach, unlike those used so far, does not require the creation of dictionaries, takes into account the broad context of words and their meaning in financial texts, eliminates the problem of ambiguity of words in various contexts, does not require manual labelling of data and is free from the subjective assessment of the researcher. The sentiment of financial texts in the meaning presented in this paper is directly related to the market reaction to the information contained in these texts. For texts belonging to one of the two classes (positive or negative) with the highest probability the BERT model gives the results of predictions with a precision level of 62.38% for the positive class and 55% for the negative class. The results at this level can be used in event study, market efficiency research, investment strategy development or support of investment analysts using fundamental analysis.

Keywords: fundamental analysis supported by machine learning; financial texts sentiment analysis; natural language processing in finance.

1. Introduction

An important role in the fundamental analysis is played by the acquisition and analysis of various types of information about the company. Text documents are an increasingly important source of this information. An example often cited in the literature are records from press conferences with the participation of companies' management and optional attachments to periodic and current reports (Palepu K., G., Healy, P., M. 2013). In addition, text data sources popular among analysts include: annual and quarterly reports with accompanying press releases, press articles, analyst reports and social media (El-Haj M. et al. 2019). The list of scientific publications in the field of finance divided into categories based on the analyzed text data sources is presented in Table 1. Research papers in this area concern mainly the American market.

Researches in the area of financial text analysis are based on sentiment analysis. Text sentiment is a concept taken from the natural language processing (NLP) literature (Jurafsky D., Martin J.H. 2020). It is used to classify texts and reflects the author's positive or negative orientation towards a given object. It can be defined as a measure of the extent to which the texts are positive or negative. In the case of stock exchange announcements, positive texts are understood as information that has a positive impact on the company's value. Negative texts are those that contain information that has a negative impact on the company's value. In some publications, in a similar sense to sentiment, the term "tone" of statement is used (Kearney C., Liu, S. 2014).

Table 1. The table shows the types of text sources analyzed in selected research works. SEC (U.S. Securities and Exchange Commission) symbols corresponding to the reports from which the data are derived are given in parentheses. Adapted from Kearney C., Liu, S. (2014) and supplemented with the latest publications.

	The type of text data	Research literature
Primary sources	Annual reports (K-10)	(Li, F. 2006) (Loughran, T., McDonald, B. 2011) (Jagadeesh, N., Wu, A.D. 2012)
	Management Discussion and Analysis - MD&A included in annual and quarterly reports, (10-Q and 10-K)	(Feldman, R. et al.2008) (Davis, A. K., Tama-Sweet, I. 2011)
	Information about initial public offerings (S-1)	(Loughran, T., Mcdonald, B. 2013)
	IPO prospectuses	(Jegadeesh, N., Wu, A.D. 2012) (Ferris, S. P. et al. 2013)
	Earnings press releases	(Henry, E. 2006),(Henry, E. 2008) (Henry, E., Leone, A. J. 2009) (Doran, J. S. et al. 2010) (Davis, A. K., Tama-Sweet, I. 2011) (Demers, E., Vega, C. 2011)
	Earnings conference calls	(Davis, A. K. et al. 2012) (Larker, D., Zakolyukina, A. 2012) (Price, S. M. et al. 2012) (Borochina, P. A. et al. 2017) (Davis, A. K. et al. 2015)
	Restatements of financial reports	(Durnev, A., Mangen, C. 2011)
	Analyst reports	(Huang, A. et al. 2013)
	Other based on variety of primary sources	(Rogers, J. L. et al. 2011)
	Information on changes in the balance of ownership (13D)	(Aydogdu M. et al. 2020)
	Information on changes in auditors of companies	(Holowczak R. et al. 2018)
Secondary sources	**Media** — Press and news services, such as Wall Street Journal, Dow Jones News Service, The New York Times, The Financial Times, The Times, The Guardian, Mirror, Thomson Reuters	(Cowles, A. 1933) (Tetlock, P. C. 2007) (Tetlock, P. C. et al. 2008) (Engelberg, J. 2008) (Sinha, N. R. 2010) (Garcia, D. 2014) (Carretta, A. et al.2011) (Engelberg, J. et al. 2012) (Ferguson, N.J. et al. 2015) (Buehlmaier, M. M. M. 2013) (Liu, B., McConnell, J. J. 2013)
	Media — Stock message boards	(Antweiler, W., Frank, M. Z. 2004) (Das, S. R., Chen, M. Y. 2007)
	Internet and social media — Twitter	(Bollen, J. et al. 2011) (Bartov, E. et al. 2018) (Sun, A. et al. 2016)

Two groups of methods are usually used to determine the measures of sentiment in financial texts - methods based on dictionaries or machine learning. Dictionary methods are most often used by researchers (Table 2). They are often referred to in the literature as 'bags-of-words' models. Text documents are treated here as a set of words that are assigned by researchers, on the basis of predefined dictionaries, to various categories (e.g. negative category and positive category). Determining the sentiment of a text in this method involves the calculation of an integrated indicator, usually based on the number of words belonging to each category. Formula 1 is a record of an example sentiment indicator.

$$SEN = \frac{WORDS_{POS} - WORDS_{NEG}}{WORDS_{ALL}} \quad (1)$$

where,

SEN – sentiment indicator,

$WORDS_{POS}$ –   number of positive words in the text,

$WORDS_{NEG}$ – number of negative words in the text,

$WORDS_{ALL}$ – sum of the number of positive and negative words (Henry, E. 2008).

The process of text sentiment analysis using the dictionary method can be divided into the following stages:

1.   Selecting the type of financial texts to be studied.
2.   Acquisition and preparation of text dataset.
3.   Dictionary selection.
4.   Design of the sentiment indicator.
5.   Determining the sentiment based on the value of the indicator.

Methods using ML (machine learning) models are less popular in the financial literature, but their importance is growing (Li, F. 2010), (Aydogdu M. et al. 2020).   Determining the sentiment within this group of methods has the following course:

1.   Selecting the type of financial texts to be studied.
2.   Preparation of training and test data - obtaining a large amount of text data of a given type and labelling them, i.e. manually assigning the sentiment value.
3.   Selection and preparation of the ML model.
4.   Training the model, i.e. calculating its parameters.
5.   Testing the model on a predefined dataset.
6.   Determining sentiment using a trained model.

The sentiment determined by the methods described above can then be applied to further research, such as: studying the effect of sentiment on market value and volatility of stocks, future income, profits or cash-flow, testing the informational value beyond the numerical information accompanying the text, testing the relationship sentiment of text information with the shortcomings of financial statements, event study. Both the dictionary and ML methods described above have some drawbacks. The dictionary methods omit the meaning of words and their wider and varied context, which is sometimes crucial for understanding the tone of given sentences. This largely limits the proper analysis of text documents. In addition, the dictionary approach encounters the problems of ambiguity of various words, especially within the context related to the issue under study. For example, the word "growth" may have a positive meaning in terms of the company's profit, but negative in terms of the number of complaints.   As in the case of dictionary methods, also in the case of some methods in the area of machine learning, used in financial literature, the mutual contextual connections between words in sentences are not taken into account. Naive Bayesian classifiers of the text are directly based on the "naive" assumption about the independence between the probabilities of the occurrence of particular words in sentences. The use of deep neural networks such as LSTM partially eliminates the problems resulting from taking into account the wider context of words in the sentence. However ML methods require a lot of work to label tens of thousands of text data.

Both groups of methods in determining sentiment are largely based on the subjective opinion of the researcher. In the case of dictionary methods, when constructing dictionaries and assigning words to individual categories, and in the case of ML during labelling training datasets, when the researcher or his team, based on their knowledge and experience, determines the degree of positive or negative tone of the text.

In this paper an alternative approach to the sentiment analysis of financial texts is proposed.   It is presented on the example of texts from press releases concerning the financial results of companies listed on American stock exchanges. These texts are linked to their impact on the abnormal return. The proposed solution does not require the creation of dictionaries, takes into account the broad context of words and their meaning in financial texts, eliminates the problem of ambiguity of words in various contexts, does not require manual

labelling, and is free from the subjective assessment of the researcher. It is based on the fact that the sentiment of press releases about financial performance influences the market response as measured by the cumulative abnormal return (CAR). This influence has been proven in many scientific publications (Table 2). So it is assumed that CAR can be used to automatically label text data. In this sense, a positive abnormal return value means that the text of the related press release has a positive sentiment and a negative value indicates a negative sentiment. This type of labelling requires significantly less work and is objective. The positive or negative sentiment in this case does not have to be the same as the one that would be subjectively marked by the researcher. A large set of text data, labelled in this way, is then used to train a deep neural networks model - BERT. This model takes into account not only the presence of words in texts, but also their place in sentences and the broad context. The sentiment determination process presented here follows the same course as that indicated above for the other ML models, with the text data not being labelled manually, but automatically using a predefined CAR-based measure.

Table 2. The table presents the leading scientific articles on the market response to the texts of earnings press releases and earnings conference calls. In addition to the publications and sources of text data, the period covered by the study, the method of text content analysis and the model used to study the relationship between the content of the messages and the market response measured by the cumulative abnormal return (CAR) are also given. The last two columns contain data about the width of the event window and a summary of the results.

Scientific publication	Study period	Source of text data	Content analysis methods	Models	Event window width	Market response
(Henry, E. 2008)	1998 – 2002	Earnings press releases	Dictionary based (Diction5.0)	Linear regression, event-study	3 days	The tone of press releases influences the market response as measured by the cumulative abnormal return (CAR)
(Henry, E., Leone, A. J. 2009)	2004 – 2006	Earnings press releases	Dictionary based (Henry DICTION, GI/Harvard)	Linear regression, event-study	3 days	The tone of press releases influences the market response as measured by the cumulative abnormal return (CAR), with stronger negative tone influence.
(Doran et al. 2010)	2004 – 2007	Earnings conference calls	Dictionary based (Henry DICTION, GI, Henry)	Linear regression, event-study	2 days, 9 days, 21 days	The tone of conference calls has an impact on the market response measured by the cumulative abnormal return (CAR) - significant in the 2-day range
(Davis et al. 2011)	1998 – 2003	Earnings press releases	Dictionary based (DICTION)	Linear regression, event-study	3 days	The language of press releases influences the market response as measured by the cumulative abnormal return (CAR)
(Demers, E., Vega, C. 2011)	1998 – 2006	Earnings press releases	Dictionary based (DICTION, GI/Harvard, LM)	Linear regression, event-study	3 days	The optimism expressed by management in the press releases has an informative content that is valued by the market (CAR)
(Price, S. M. et al. 2012)	2004 – 2007	Earnings conference calls	Dictionary based (GI/Harvard Henry)	Linear regression, event-study	3 days, 59 days	The language of conference calls influences the market response as measured by the additional rate of return (CAR)

2. Materials and Methods

2.1. Data sources and transformation of data into a form suitable for the ML model

U.S. public companies are required to publish the information required by law in electronic form through the EDGAR (Electronic Data Gathering, Analysis, and Retrieval System) operated by the U.S. Securities and Exchange Commission (SEC). The EDGAR system processes approximately 3,000 electronic publications per day and makes publicly available 3 petabytes of data per year. Access to the public database of the EDGAR system is unlimited and free of charge.   Pursuant to the provisions of the American securities law of 1934 (Securities Exchange Act of 1934 section 13 or 15 (d)), companies are required, in addition to annual reports (form 10-K) and quarterly reports (form 10-Q), also to publish current reports (form 8-K). Current reports are submitted in case of the event or circumstances that the shareholders should learn about. The form 8-K contain 9 sections with a total of 31 items such as: entry into a material definitive agreement, declaration of bankruptcy or receivership, results of operations and financial condition, unregistered sales of equity securities, departure of directors or certain officers and others. Section 9 contains certain financial statements and list the exhibits that it has filed as part of the 8-K form. In most cases, companies have 4 days from the occurrence of the event to fulfil their obligation to publish the current report.   The classification of the scope of information that should be disclosed in the 8-K filings when it occurs is presented in Table 3.

Table 3. Categories of events, the occurrence of which should result in the publication of relevant information in the current report 8-K.

Categories of events	Scope of information
Registrant's business and operations	Entry and termination of material definitive agreement, bankruptcy or receivership, reporting of shutdowns and patterns of violations in mines.
Financial Information	Acquisition or disposition of assets, results of operations and financial condition, creation and change of a balance sheet or off-balance sheet liability, costs associated with exit or disposal activities, material impairments.
Securities and trading markets	Issues concerning delisting for any class of the registrant's common equity, unregistered sales of equity securities, modification to rights of security holders.
Matters related to accountants and financial statements	Changes in company's certifying accountant, non-reliance on previously issued financial statements or a related audit report or completed interim review.
Corporate governance and management	Changes in control of registrant, changes in management stuff, amendments to articles of incorporation or bylaws, change in fiscal year, temporary suspension of trading under registrant's employee benefit plans, amendments to the registrant's code of ethics, change in shell company status, submission of matters to a vote of security holders, shareholder director nominations.
Asset-backed securities	Informational and computational material, change of servicer or trustee, change in credit or other external support, failure in securities distribution.
Fair disclosure regulation	Disclosure of any information that has been shared with other certain individuals or entities.
Other Events	Any events, with respect to which information is not otherwise called for by 8-K form, that the registrant deems of importance to security holders.
Financial statements and exhibits	Pro forma financial information and **exhibits**, financial statements of businesses or funds acquired, pro forma financial information, shell company transactions, other exhibits.

Often, companies announce their quarterly and annual results at conference calls immediately before or simultaneously with the publication of the report. In such cases, the content presented at the conference call and the summary of the financial reports constitute an appendix marked as 'EXHIBIT 99' to the 8-K form. This exhibit may also contain additional information that is not disclosed under other types of exhibits.

The text research data comes from the 'EXHIBIT 99' appendices to the 8-K current reports published by the companies included in the S&P 500 index. All reports were published in the EDGAR system. If a given report had an EXHIBIT 99 attachment, its text content as well as the date and exact time of publication were extracted. Text data has been "cleaned", i.e. they have been deprived of irrelevant data, e.g. contact information, redundant spaces, references to the attachments, etc. Moreover, due to the available computing power, the texts were shortened to the initial 256 words. The market data comes from the INTRINIO service and includes adjusted (after taking into account dividends pay outs and splits) daily stocks prices of the companies covered by the study and the value of the S&P 500 index. Both textual and financial data cover the period from 02/06/2014 to 31/12/2019. For each current report, the cumulative abnormal return (CAR) is calculated, defined as the difference between the return on shares minus the return on the S & P500 index over a 9-day period constituting the so-called "event window" - Formula 2. The event window starts 4 days before the report publication date and ends 4 days

after that date. When determining the width of the window, the values adopted in other publications, ranging from 2 to 59 days, were taken into account (Table 2). It also takes into account the fact that in most cases companies have 4 days from the occurrence of the event to fulfil the obligation to publish the current 8-K report.

$$CAR_t^i = R_t^i - IR_t \qquad\qquad\qquad (2)$$

where,

t – event window width of 9 days (period starting 4 days before the publication of the report and ending 4 days after that date)

CAR_t^i – stock i' cumulative abnormal return at period t;

R_t^i – stock i' return at period t;

IR_t – S&P 500 index return at period t.

It is common practice to use a stock index return as the benchmark for calculating CAR. This is the so-called naive model. Other benchmarks may be for example calculated from: Sharpe's single-index model (1963), multiple factor models or CAPM (capital asset pricing model). (Klinger, D., Gurevich, G. 2014).

The abnormal return calculated according to Formula 2 were used to assign the impact of the publication on share prices to two classes marked with appropriate labels: POSITIVE for CAR ≥ 0 and NEGATIVE for CAR <0. Finally the data set consisted of 6,435 samples including the text of the appendix 'EXHIBIT 99' to the 8-K current report and a class label indicating the category of the share price change. An illustrative fragment of the training set is presented in Table 4.

Table 4. Illustrative fragment of the training set. The first column contains the text extracted from the appendix 'EXHIBIT 99' to the 8-K current report. The second column contains the labels of the sentiment classes calculated for the text. The first text concerns Akamai Technologies, Inc. - American provider of cloud services and the second concerns company from the transport sector - C.H. Robinson Worldwide, Inc.

Text	Class label
„(nasdaq: akam), the world's largest and most trusted cloud delivery platform, today reported financial results for the fourth quarter and full-year ended december 31, 2018. "we were very pleased with our strong finish to the year. both revenue and earnings exceeded our expectations due to the very rapid growth of our cloud security business, robust seasonal traffic and our continued focus on operational excellence," said dr. tom leighton, ceo of akamai. "as a result, we achieved our fifth consecutive quarter of non-gaap operating margin improvement, and we are well on our way to achieving our 30% margin goal in 2020 …"	POSITIVE (1)
„(nasdaq: chrw) today reported financial results for the quarter ended september 30, 2019. "the third quarter provided challenges in both our north american surface transportation and global forwarding segments. our net revenues, operating income, and eps results finished below our long-term expectations. we anticipated an aggressive industry pricing environment coming into the second half of this year driven by excess capacity and softening demand and knew we faced difficult comparisons versus our strong double-digit net revenue growth in the second half of last year. our results were negatively impacted by truckload margin compression in north america," said bob biesterfeld, chief executive officer …"	NEGATIVE (0)

The dataset was randomly split into two subsets. Training dataset of 5,148 samples (80%) and test dataset of 1,287 samples (20%). The data covers the period from 02-06-2014 to 31-12-2018. A set of additional test data, which is not involved in the model training process was also prepared. It includes 1831 samples from 01/01/2019 to 31/12/2019 and will be used for the final verification of the model.

Then the data was transformed into a form that can be loaded into the model. This process includes, among other things, the "tokenization" of texts which transforms words into numbers. Finally the data is converted to files in the TFRecord binary format used by the TensorFlow library created in the Python programming language.

2.2. Basic features of the BERT model used for sentiment analysis

BERT (Bidirectional Encoder Representations from Transformers) (Devlin, J. et al. 2018) is a natural language processing model built of deep neural networks proposed by the Google AI Language team in 2019. It performs exceptionally well in "understanding" natural language compared to other general-purpose NLP models. It is the first unsupervised and bidirectional NLP model. No supervision means there is no need to use labelled data to train it. The model's bidirectional nature means that the vector representation of a word it generates depends on other words in the sentence, both before and after the given

word. The built-in attention mechanism is a very important element of the model. Thanks to this, it takes into account the broad context of the words in the text. Google has released both the model's source code and pre-trained models. A text data corpus from Wikipedia and BookCorpus was used to train them. The BERT model can be easily adapted to many types of NLP tasks, such as classifying texts or questions and answering. The adaptation process consists in fine tuning the weights of the model during additional training on the labelled data. At the stage of training the model (pre-train), he acquires 'knowledge' about the structure and relations within a given natural language, and at the stage of fine-tuning about a specific domain related to a given task. Models provided by Google can process sentences containing a maximum of 512 words. In this paper, the sentence length is limited to 256 words The diagram of the model's operation is presented in Figure 1. The input data in the form of words is marked in pink. Words are converted into 768 dimensional vectors. Then vectors representing the position of the word in the sentence are added to them (yellow). The output (green) is the new word representations. The first vector of the model marked as [CLS] is used to perform classification tasks. A full description of the model can be found in (Devlin, J. et al.2018).

Figure 1. Diagram showing the basic functional elements of the BERT model (Devlin, J. et al.2018).
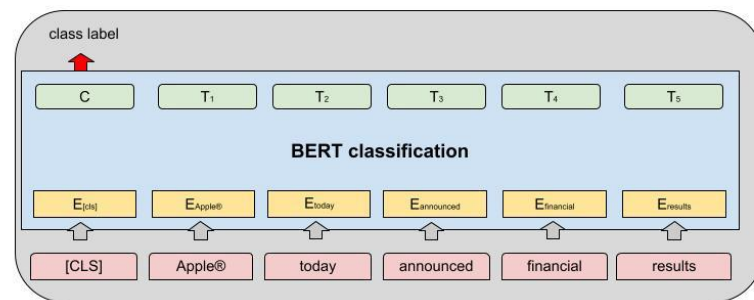
The BERT model was chosen because of the following features:

1. It is currently one of the best performing models in natural language processing.
2. Takes into account the order of words in the text.
3. Through the attention mechanism, it takes into account the broader context of words in the text.
4. It can be easily adapted to different categories of NLP tasks.

In classification task, the output data from BERT model is a vector of numerical values with a dimension consistent with the number of classes. Then these values are transformed using the softmax function (also known as the normalized exponential function) into the vector of probabilities that the sample belongs to each class.

3. Results

The data prepared as described in Section 2 above was used to fine-tune the BERT model. It should be noted that fine-tuning the model is simply training it further with the difference that labelled data and subject-specific texts are used. Fine-tuning requires a lot of computing power, although less than pre-training. In this work, the basic BERT model (BERT-Base) was used, in which over 109 million weights of the neural network require tuning. The model was adapted to the task of classifying the text into two classes (POSITIVE, NEGATIVE). The input texts in each example include a maximum of 256 words

converted to the corresponding numeric form. The model was tuned over 6 training epochs. In each epoch, the model optimizes its parameters based on all training samples grouped into mini-batches of 4 samples each. For a training set of 5148 samples, each epoch consists of 5148/4 = 1287 steps. At each step, the model calculates a cost function for the current mini-group. Then, using the back propagation algorithm, the gradients of this function are calculated as well as new weights of the neural network. After each epoch, the model is evaluated on a test dataset. It involves a partially trained model making predictions and comparing them with the labels assigned to the test data. Finally a measure is calculated for the entire set, on the basis of which the effectiveness of the model can be assessed at a given stage.

Precision was used to evaluate the model - a measure commonly used in classification problems - Formula 3.

$$A = 100\% * \left(\frac{PP}{PP+FP}\right) \hspace{3cm} (3)$$

where,
A – precision;
PP – the number of true positives;
FP – the number of false positives.

This measure indicates how often the model's predictions match the labels.

The minimum value of the cost, calculated on the test set, was achieved by the model in the second epoch of training. Then the amount of cost increases more and more rapidly, which is probably caused by the model overfitting - Figure 2.

Figure 2. The figure shows the cost (ordinate axis) calculated on the test data in subsequent training epochs (horizontal axis). It reaches its minimum in the second training epoch. Then it increases due to model overfitting.
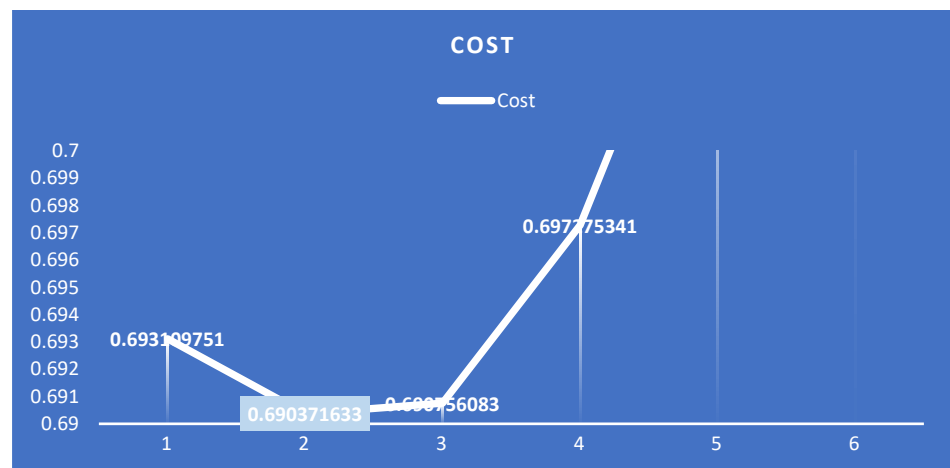
The precision for the second epoch in which the cost is the lowest and the model shows no signs of overfitting is 52.68% – Figure 3

Figure 3. The figure shows the value of precision   (vertical axis) in subsequent training periods (horizontal axis). For the second training epoch it is 0.526806533, i.e. 52.68% rounded.
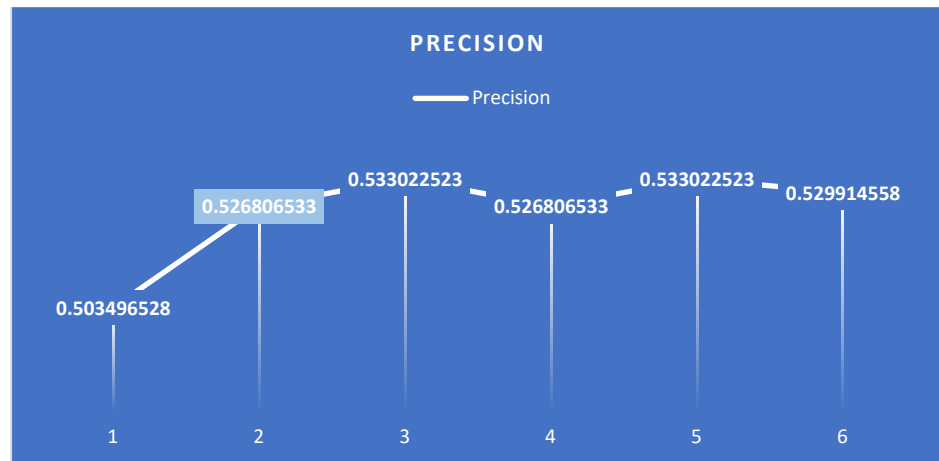
For further calculations, a model with weight values that were achieved after the 2nd training epoch was used. This model was used to calculate the probability of belonging to the POSITIVE or NEGATIVE categories for 1,831 samples from additional test dataset. From this set 101 samples were selected for which the probability of belonging to the POSITIVE class is the highest and 100 samples for which the probability of belonging to the NEGATIVE class is the highest. Then, a measure of precision was calculated for the sets of samples distinguished in this way. Result was 62.38% for the positive class and 55% for the negative class.

Sample text that has been considered POSITIVE by the model with high probability:

"apple reports fourth quarter results services revenue reaches all-time high of $12.5 billion eps sets new fourth quarter record of $3.03 cupertino, california — october 30, 2019 — apple® today announced financial results for its fiscal 2019 fourth quarter ended september 28, 2019. the company posted quarterly revenue of $64 billion, an increase of 2 percent from the year-ago quarter, and quarterly earnings per diluted share of $3.03, up 4 percent. international sales accounted for 60 percent of the quarter's revenue. "we concluded a groundbreaking fiscal 2019 with our highest q4 revenue ever, fueled by accelerating growth from services, wearables and ipad," said tim cook, apple's ceo. " with customers and reviewers raving about the new generation of iphones, today's debut of new, noise-cancelling airpods pro, the hotly-anticipated arrival of apple tv+ just two days away …"

Sample text that has been considered NEGATIVE by the model with high probability:

„(nasdaq: aal) today reported its first-quarter 2019 results, including these highlights: "we want to thank our 130,000 team members for the outstanding job they did to take care of our customers, despite the challenges with our fleet during the quarter. their hard work led american to record revenue performance under difficult operating conditions," said chairman and ceo doug parker. "as we progress toward the busy summer travel period, demand for our product remains strong. however, our near-term earnings forecast has been affected by the grounding of our boeing 737 max fleet, which we have removed from scheduled flying through aug. 19. we presently estimate the grounding of the 737 max will impact our 2019 pre-tax earnings by approximately $350 million. with the recent run-up in oil prices, fuel expenses for the year are also expected to be approximately $650 million higher than we forecast just three months ago. …"

4. Discussion

The BERT model used in the study can determine the sentiment of financial texts understood as a measure of its impact on the abnormal return on shares. The model gives

52.68% precision of predictions in the case of the test data set, i.e. 1.68% more than the percentage level of a larger category. For samples defined as positive or negative with the highest probability level, the measure of precision is much higher and amounts to even 62.38% for the POSITIVE category. Although the results of classification using the BERT model for more classic NLP tasks are usually higher, it can be concluded that they should not be compared with the issue described in this paper. It should be noted that even a very positive text, as read by a financial analyst, does not necessarily produce a positive market response. Similarly very negative texts of current reports will not always cause a fall in share prices on the stock exchange. In this case, the model must sometimes look for very subtle meanings in the texts in order to achieve satisfactory results.

In addition, the model created in this study does not require the creation of dictionaries, takes into account the broad context of words and their meaning in financial texts, eliminates the problem of ambiguity of words in various contexts, does not require manual labelling, and the training process is free from the subjective assessment of the researcher. The examples of texts given in the previous section that are defined as strongly positive or strongly negative would probably be identified in the same way by most professionals using fundamental analysis.

The method presented here can be used in the following research and practical areas:
1. Event study.
2. Market efficiency research.
3. Investment strategies.
4. Support for investment analysts using fundamental analysis.

The model may also be adapted to characterize the sentiment of financial texts in languages other than English.

References

Antweiler, W., Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. Journal of Finance, 59, 1259–1294

Aydogdu M, Saraoglu H, Louton D. (2020). Using long short-term memory neural networks to analyze SEC 13D filings: A recipe for human and machine interaction. Intell Sys Acc Fin Mgmt. 2020;1–11. https://doi.org/10.1002/isaf.1464

Bartov, E., Faurel, L., Mohanram, P. (2018). Can Twitter help predict firm-level earnings and stock returns? The Accounting Review, 93(3), 25–57.

Bollen, J., H. Mao, and X. Zeng (2011). Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8.

Borochina, P. A., Ciconb, J. E., DeLislec, R. J., McKay Priced, S. (2017). The effects of conference call tones on market perceptions of value uncertainty. Journal of Financial Markets, Forthcoming. SSRN: https://ssrn.com/abstract=2579907 or http://dx.doi.org/10.2139/ssrn.2579907

Buehlmaier, M. M. M. (2013). The role of the media in takeovers: Theory and evidence. SSRN: http://ssrn.com/abstract=1673164

Carretta, A., Farina, V., Graziano, E. A., Reale, M. (2011). Does investor attention influence stock market activity? The case of spin-off deals. SSRN:http://ssrn.com/abstract=1930861

Cowles, A. (1933). Can stock market forecasters forecast? Econometrica 1(3), 309–324.   http://www.jstor.org/stable/1907042

Das, S. R., Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management Science, 53, 1375–1388.

Davis, A. K., Tama-Sweet, I. (2011). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. Available at SSRN. http://ssrn.com/abstract=1866369

Davis, A.K., Piger, J.M., Sedor, L.M. (2011). Beyond the numbers: Measuring the information content of earnings press release language. AAA 2008 Financial Accounting and Reporting Section (FARS) Paper (Available at SSRN: http://ssrn.com/abstract=875399).

Davis, A. K., Ge, W., Matsumoto, D. A., Zhang, J. L. (2012). The effect of managerial 'style' on the tone of earnings conference calls. CAAA Annual Conference 2012 (SSRN:http://ssrn.com/abstract=1982259).

Davis, A. K., Ge, W., Matsumoto, D., Zhang J. L. (2015). The Effect of Manager-Specific Optimism on the Tone of Earnings Conference Calls. Review of Accounting Studies 20:639–73.

Demers, E., Vega, C. (2011). Linguistic tone in earnings announcements: News or noise? FRB International Finance Discussion Paper 951 (SSRN: http://ssrn.com/abstract=1152326)

Devlin, J., Chang, MW., Lee, K., Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language under-standing, arXiv:1810.04805 [cs.CL]. https://github.com/google-research/bert

Doran, J. S., Peterson, D. R., Price, S. M. (2010). Earnings conference call content and stock price: The case of REITs. Journal of Real Estate Finance and Economics, 45(2), 402–434.

Durnev, A., Mangen, C. (2011). The real effects of disclosure tone: Evidence from restatements. Available at SSRN. http://ssrn.com/abstract=1650003

Engelberg, J. (2008). Costly information processing: Evidence from information announcements. AFA 2009 San Francisco Meetings Paper (SSRN: http://ssrn.com/abstract=1107998).

Engelberg, J., Reed, A. V., Ringgenber, M. C. (2012). How are shorts informed? Short sellers, news, and information processing. Journal of Financial Economics, 105(2), 260–278.

Feldman, R., Govindaraj, S., Livnat, J., Segal, B. (2008). The incremental information content of tone change in management discussion and analysis. Available at SSRN. http://ssrn.com/abstract=1126962

Ferguson, N.J. ,Philip, D., Lam, H. Y. T., Guo, J. M. (2015). Media content and stock returns: The predictive power of press. Multinational Finance Journal, 2015, vol. 19, no. 1, pp. 1–31.

Ferris, S. P., Hao, G. Q., Liao, M. (2013). The effect of issuer conservatism on IPO pricing and performance. Review of Finance, 7(3), 993–1027.

Garcia, D. (2014). Sentiment during recessions. Journal of Finance (SSRN: http://ssrn.com/abstract=1571101).

Henry, E. (2006). Market reaction to verbal components of earnings press releases: Event study using a predictive algorithm. Journal of Emerging Technologies in Accounting, 3, 1–19.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? The Journal of Business Communication, 45(4), 363–407

Henry, E., Leone, A. J. (2009). Measuring qualitative information in capital markets research. Available at SSRN. http://ssrn.com/abstract=1470807

Holowczak, R., Louton, D., Saraoglu, H. et al. (2018). Testing market response to auditor change filings: A comparison of machine learning classifiers. https://doi.org/10.1016/j.jfds.2018.08.001

Huang, A., Zang, A. Y., Zheng, R. (2014). Large sample evidence on the informativeness of text in analyst reports. The Accounting Review American Accounting Association, Vol. 89, No. 6, 2014 pp. 2151–2180, DOI: 10.2308/accr-50833

Jegadeesh, N., Wu, A.D. (2012). Word power: A new approach for content analysis. AFA 2012 Chicago Meetings Paper (Available at SSRN: http://ssrn.com/abstract=1787273).

Jurafsky D., Martin J.H., (2020). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Stanford University 2020.

Kearney, C., Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171–185.

Klinger, D., Gurevich, G. (2014). Event studies for financial research: a comprehensive guide. Palgrave Macmillan.

Larker, D., Zakolyukina, A. (2012). Detecting deceptive discussions in conference calls. Journal of Accounting Research, 50(2), 495–540.

Li, F. (2006). Do stock market investors understand the risk sentiment of corporate annual reports? Available at SSRN. http://ssrn.com/abstract=898181

Li, F. (2010). The information content of forward-looking statements in corporate filings - A Naive Bayesian machine learning algorithm approach. Journal of Accounting Research, 48, 1049–1102.

Liu, B., McConnell, J. J. (2013). The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? SSRN:http://ssrn.com/abstract=2004567

Loughran, T., McDonald, B. (2011). When is a liability not a liability? Journal of Finance, 66, 35–65.

Loughran, T., Mcdonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and Form S-1 language. Journal of Financial Economics, 109(2), 307–326.

Price, S. M., Doran, J. S., Peterson, D. R., Bliss, B.A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. Journal of Banking and Finance, 36(4), 992–1011.

Rogers, J. L., Van Buskirk, A., Zechman, S. L. C. (2011). Disclosure tone and shareholder litigation. Accounting Review, 86(6), 2011.

Sinha, N. R. (2010). Underreaction to news in the US stock market. Available at SSRN. http://ssrn.com/abstract=1572614

Sun, A., Lachanski, M. Fabozzi, F.J., (2016) Trade the Tweet: Social Media Text Mining and Sparse Matrix Factorization for Stock Market Prediction. International Review of Financial Analysis. doi: 10.1016/j.irfa.2016.10.009

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62, 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. Journal of Finance, 63, 1437–1467.