BIG DATA ANALYTICS AND FIRM PERFORMANCE: A TEXT MINING APPROACH

Mihai BOGDAN[a]**, Anca BORZA[b]

[a,b] Babeş-Bolyai University, Cluj-Napoca, Romania

ABSTRACT

Big data analytics, used to be a state-of-the-art process, now it is a must. Nowadays, firms promote this as a standard operating procedure. It is not only the luxury of the large multinationals, but we see, that also local companies from the Romanian economy start to present this in their annual financial reports. Numerous studies tried to measure and prove the positive impact of big data analytics upon firm's performance. In the current study we do not aim to do the same. Instead, we propose to review all the studies on this topic, which we could find, and to test a common practice in the field of big data analytics called sentiment analysis, which is used to assess the attitude of the writers towards their subject. Our hypothesis is that there is a strong relationship between the reported results of the studies, called effect sizes and the wording or phrasing of the articles, called sentiment score. As a methodology, we used Publish or Perish which is an academic "search engine" to search all the articles on the big data analytics – firm performance topic available on Google Scholar between 2010-2020. Our initial search found over5500 articles, books and other publications. Based on our criteria, we kept 120 articles. The main finding is that there is no correlation between the effect sizes of the articles and the sentiment score we obtained. Moreover, we found that on average, based on the wording used, the studies tend to overstate their results.

KEYWORDS: *big data analytics, firm performance, sentiment analysis*

1. INTRODUCTION

Each resource is valuable based on the purpose and the scope for which is being used. Big data analytics on its own can bring no value to an organization. Its sole utility is determined based on the number of valuable insights brought to the organization, and the degree to which the decision makers of the organization base their decisions on these findings. It's a two-way relationship, in which both sides have their critical contribution.

Based on the literature we reviewed, we found two levels of big data *"integration"*. First level is a big data analytics initiative within an organization. At this stage the organization has people with the skills and the technology necessary for such extensive analyses. But all these activities are seen as a separate department or function and are usually treated as a support function. Second stage, is where big data analytics become a capability for the organization and all these is seen as a strategic resource. In this stage, the organization has in place the resources, the technical skills but additionally, also the management skills, the right culture andan organizational learning *"mindset"*. The second stage can put the organization on a favorable spot, called competitive advantage. As Jeble et al. (2018) wrote, while resources can be procured on the market, capabilities (e.g. data driven decision, organizational learning) have to be developed *"in house"*.

---

* Corresponding author. E-mail address: mihai.bogdan@econ.ubbcluj.ro

2. LITERATURE REVIEW

In this section we aim to present the relationship between the forms big data analytics and firm performance. As a summary, in the literature that we reviewed, big data analytics appears in many forms such as: big data analytics capability, information processing capability, analytical strategy, data driven environment, big data investments announcements, ad-hoc/ routine use of big data, customer analytics and Internet of Things. On a similar note, the positive effects of big data analytics, in other words, firm's performance is viewed in multiple forms as well: financial performance, competitive advantage, operational performance, customer satisfaction, organizational agility, market valuation (stock price), decision making effectiveness, innovation capability, supply chain performance, visibility and coordination, organizational learning social and environmental performance.

2.1 The path from data to big data analytics (BDA) to big data analytics capability (BDAC)

Adoption of big data analytics (BDA) would involve deploying operation towards descriptive analytics, predictive analytics, and prescriptive analytics together with data acquisition and processing capabilities (Aydiner et al., 2019). Popovič et al. (2019) identify the routine use and the innovative use of BDA. Routine use indicates employees using data analytics in standardized and procedural method. Innovative use means moving beyond the daily routine and discovering new ways of analyzing data and ultimately new insights. A partially similar approach is followed by Hosoya and Kamioka (2018) where ad-hoc use of big data analytics consists of: data collection and analysis, interpretation (data given meaning) and learning (action taking). Ghasemaghaei (2018) evaluates BDA in terms of: tools sophistication, BDA utilization and employees' analytical skills. Also, for Thirathon et al. (2017), the level of BDA sophistication is measured in terms of analytic tools and analytics methods.

Further, in order of big data analytics to be effective, a data driven environment has to be nurtured, and managers have to adopt a data-driven decision-making style (Cao et al., 2015). This will create a data-driven culture, where data and data-insights become part of the value system of the organization (Duan et al., 2020).

The big data analytics capability (BDAC) is approached using classical views such as resource-based theory (RBT) or under dynamic capability perspective. According to Barney (1991), in the resource-based view, the sustained competitive advantage is generated if the organization manages to assemble a set of resources which are valuable, difficult to imitate, and non-substitutable. In the context of BDAC, Gupta and George (2016) propose a model with the following resources: tangible (data, technology, and basic resources in the form of time and investments), human (management skills and technical skills), and intangible (data-driven culture and intensity of organizational learning). Akter et al. (2016) designed a similar model where BDAC consists in a set of sub-capabilities which are further detailed in other processes or resources as it follows: BDA management capability (planning, investment, coordination, and control), BDA technology capability (connectivity, compatibility, and modularity) and BDA talent capability (technology management knowledge, technical knowledge, business knowledge and relational knowledge).According to the dynamic capability perspective theory, an organization integrates external and internal competences in order to develop new processes or products and to respond to the constantly changing environment (Singh & Singh, 2019; Mikalef et al., 2020). When it comes to big data analytics capability the differentiators are not the resources themselves, but rather how they are managed in order to obtain a competitive advantage (Hao et al., 2019). Another framework under which BDAC is evaluated is through the knowledge-based view. In these terms, the bedrock of the competitive advantage relies in employees' skills and the processes through which this

knowledge is disseminated within the organization. In this regard, BDAC is the generator of new knowledge by all the insights obtained (Ghasemaghaei, 2019).[†]

Finally, no matter how advanced or complicated, big data analytics, needs to be aligned to the organization's strategy, otherwise it risks to become just another initiative meant to associate the organization with the today's buzzwords. When putting BDA in the context of leadership McAfee and Brynjolfsson (2012) beautifully put it: *"Companies succeed in the big data era not simply because they have more or better data, but because they have leadership teams that set clear goals, define what success looks like, and ask the right questions. Big data's power does not erase the need for vision or human insight."*

2.2 The many forms of performance

When it comes to positive effects of big data analytics, we will call this performance. And this element is approached from multiple perspectives. Most often encountered are the following: financial performance, operational performance, market performance, competitive advantage (strategic performance), customer satisfaction, innovation, decision making effectiveness, organizational learning, employee development, social and environmental performance.

Financial performance is one of the most classical and most encountered approach to assess performance. That can be because it is easy to be measured. Financial indicators for profitability (return on assets, return on equity, return on sales, or profit margin), for capital structure (e.g. debt to equity or financial leverage) are used by Huang et al. (2020) and Tang et al. (2018). As a similar approach, we also found, the operational performance usually treated as productivity and measured through indicators such as: inventory turnover, labor productivity, accounts receivable turnover, collections efficiency etc. (Huang et al., 2020; Tang et al., 2018).

Market performance is usually evaluated in comparison with the competition. For example, comparing how fast a firm introduces new product or services and the success rate of these, how often it enters new markets, and also in terms of market share (Wamba et al., 2017; Gupta et al., 2019).

Competitive advantage (or strategic performance), although a much subjective indicator, usually is defined as the organization's ability to obtain superior performance compared to its competitors, and if this is maintained for a longer period of time, then is called sustained competitive advantage. This is possible when the organization has access to resources or capabilities processes which are valuable, rare, difficult to imitate and do not have substitutes (Côrte-Real et al., 2017). We briefly showed with the help of literature, how big data analytics can become big data analytics capability (Dubey et al., 2019).

Customer satisfaction (together with customer relationship) is also a broad topic, but for our purpose we mention how big data analytics can help to identify the needs of the customers, customize the products or services, reduce customer acquisition costs, maintain a closer supplier-customer relationship and increase loyalty over time. (Raguseo & Vitari, 2018).

Organizational agility is seen as a dynamic capability which enables the organization to adapt to the continuously changing environment by reconfiguring resources or processes and developing new products or services. Here data analytics help environmental scanning and market research (Côrteet al., 2017; Ghasemaghaei et al., 2017). Innovation is defined on a similar note. In this direction, big data analytics appears as an enabler of innovation by harnessing tacit knowledge and enhancing the dissemination of information. All this will have a positive effect upon research and development of new products and services (Niebel et al., 2019).

[†]

Decision making effectiveness is assessed on the extent to which a decision helps to achieve the objectives of the organization. With the use of big data analytics managers can get insights closer to reality and also improve the speed of decision (Wang & Byrd, 2017).

When it comes to supply chain performance there are many definitions, but maybe the most comprehensive one is the approach from Yu et al. (2018) which conceptualize the framework of supply chain performance into four pillars: information exchange, coordination, activity integration and responsiveness across the supply chain.

Big data analytics contributes as well in the process of employees' development by nurturing a continuous learning culture, improving the level of knowledge, visibility, transparency ultimately increasing collaboration (Bag et al., 2020).

Finally, big data analytics contributes to environmental performance by reducing air emissions, reducing waste water, reducing solid wastes, decreasing consumption of harmful materials and improves the overall organization's environmental situation. In terms of social performance, we see positive effects such as: total employment, employees per enterprise, average salaries, and male vs. female employment (Dubey et al., 2019).

3. RESEARCH METHODOLOGY

Big data analytics*"ambassadors"*promise a lot of positive effects from this initiative. In our approach we tried to use a methodology that is well known in the world of data scientists called *sentiment analysis*. Kwartler (2017) defined sentiment analysis as *"the process of extracting an author's emotional intent from text."* In other words, sentiment analysis or sentiment scoring is used to assess the attitude of the writer towards a specific topic. This procedure is very often used in scoring the sentiment of the customers towards products or services extracted from the online reviews. Another example would be to estimate the attitude from news posts. For example, someone would be interested to score how a specific event made the news.

We wanted to test this methodology against all the empirical research we could find in the Google Scholar database on the topic of big data analytics and firm performance. To enhance our search and find as many entrances as possible we used Publish or Perish, which is a software that imports all the citations from an academic database, in our case Google Scholar. Given that the topic is relatively new, we set the time-frame between January 2010 and July 2020. In our search we created a searching formula to return the following topics: big data or business analytics, combined with performance or competitive advantage or value, combined with management and empirical. After that we listed all the possible combinations, which resulted in the following formula:

"big data AND performance AND empirical AND management OR big data AND competitive advantage AND empirical AND management OR big data AND value AND empirical AND management OR business analytics AND performance AND empirical AND management OR business analytics AND competitive advantage AND empirical AND management OR business analytics value AND empirical AND management."

After we run our search, we retrieved 5614 articles in our results. These were passed through a first filter where we evaluated if the article matches our criteria by reading the title or the abstract where the title would not be very clear. After this iteration cycle, we kept 526 articles. The next filter we applied was based on the content itself of the article. At this step the requirement was for an empirical research and the effect sizes had to be specified. By applying this filter, we were left with our final database of 120 articles. In *Figure 1* we show as well how our database became smaller and smaller after each iteration step.
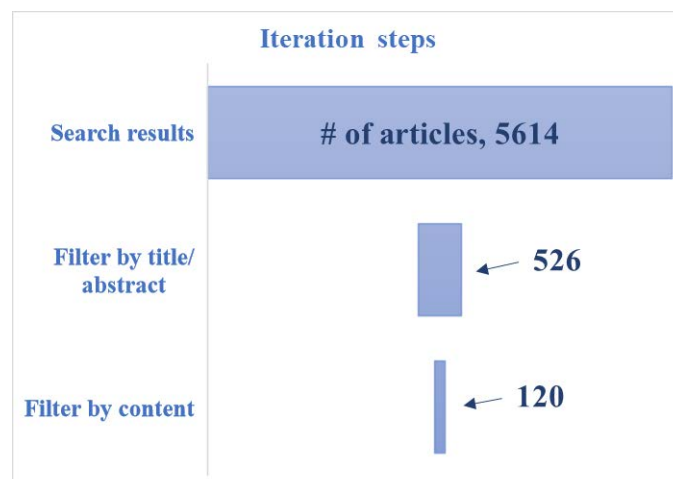
Figure 1. Iteration steps in creating the articles database for the study
Source: Own editing

As the body of text for our analysis we extracted from each article those parts where the results and conclusions were presented. We wanted to extract this specific parts only because in sections such as introduction, literature review or methodology, the tone of the text might be biased. In the first parts of the article, the attitude of the writers is influenced by the investigated literature or by their ambitious hypothesis. Because of this we extracted the text only from sections such as the following: data analysis & results, hypotheses testing, discussion, managerial contributions, theoretical implications, contribution of the study, and conclusions.

So, at the end of this extensive operation of searching, filtering and text extraction we had the final database with the text corpus from each article. The next part is the calculation of the sentiment score for each article and to run a correlation analysis with the empirical results of the articles.

4.   DATA ANALYSIS

We analyzed our data in RStudio, a data science, open source software. We used the following libraries: *dplyr*, *stringr*, *tidytext*, *tidyr*, *ggplot2* and *igraph*.When doing a sentiment analysis simplest approach would be to appraise the words of the text in terms of positive vs. negative. More depth can be applied to the analysis by involving a larger range of sentiments such as: litigious, negative, positive, superfluous and uncertainty (Silge & Robinson, 2017). Another example is proposed by Kwartler (2017) who applies the so called "*Plutchik'sweel of emotion*" which contains eight primal emotions: anger, fear, sadness, disgust, surprise, anticipation, trust and joy.
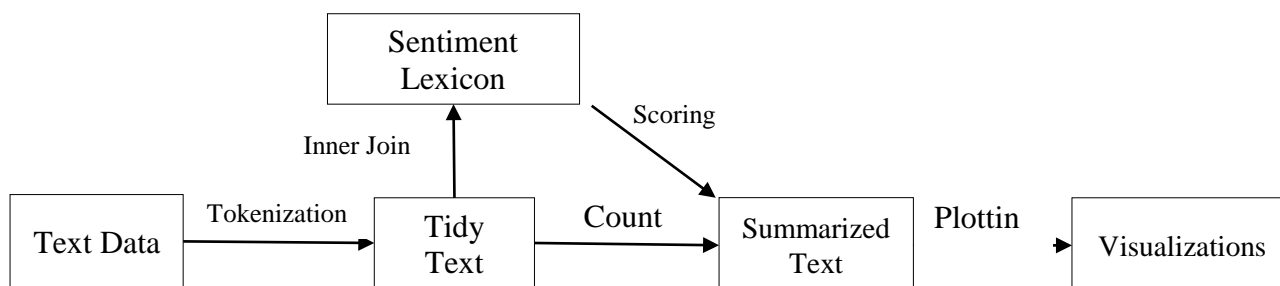

Figure 2. A flowchart of a typical text analysis that uses tidytext for sentiment analysis.
Source: adapted from Silge and Robinson (2017, p.15)

Text data is usually unstructured data, meaning it is not stored on rows and columns, but rather into sentences and paragraphs. Because of this it is necessary to be structured in order to run it through a mathematical algorithm. One of the ways in which the standardization is done is to split the text one word per row and all the upper cases are transformed into lower cases. This operation is called *tokenization*. Next *stop words* are removed from the text. These are link words such as "the", "to", "of" etc. which do not bring value to the analysis and are usually very frequent so it would load the data unnecessary. After these steps, the text would be standardized into a so-called *tidy text*. From now on, mathematical operations could be performed on it, such as counting the most frequent words, extracting the most important words from the text based on term frequency or even establishing relationships between words.

In our analysis the aim was to run a sentiment analysis for the text of each article included in our database. In order to do so, we need to run the standardized text of each article against a *sentiment lexicon*. Lexicons contain a big amount of the words in English and each have assigned a positive or negative score, or more complex emotions. Four of the most used lexicons are the following (Silge & Robinson, 2017):

- AFINN from Finn Årup Nielsen

- Bing from Bing Liu and collaborators

- NRC from Saif Mohammad and Peter Turney

- Loughran: dictionary of financial terms from Loughran and McDonald (2011)

All the lexicons, except for AFINN, label the words in terms of *positive* or *negative*. AFFIN lexicon is different in the way that it assigns a score on a range from -5 to 5 to each word. Because of this, we could not include it in our analysis.

The technique used here is that the algorithm would do an *inner join* between the text extracted from the articles and the lexicon and it will count how many words belong to each category. Based on this each article will receive a score for each of the sentiment categories: constraining, litigious, negative, positive, superfluous, uncertainty.

fter that in order to generate an overall score per article, we focused on the relationship between positive and negative applying the following formula:

$$Article\ Sentiment\ Score = \frac{postive - negative}{positive + negative}$$

In *Table 1* we display the descriptive statistics about the sentimental analysis scores for each article, when scoring against the three lexicons: Loughran, Bing, and NRC. Onaverage, we see that the most *"moderate"* one is Loughranwith an average score per article of 0.43, then for Bing lexicon the average score per article is 0.65, while when using NRC lexicon, the average score is 0.75. Now we would like to highlight here that the most trustworthy for our analysis would be Loughranbecause this one is specifically designed for financial/ economical terminology. The possible scores ranged from -1 to 1.We see that depending on the lexicon used the ranges of the scores are the following: Loughran [–0.33; 1], Bing [0.02; 1], and NRC[0.38, 0.97].

Table 1. Descriptive statistics for the sentiment scores per article based on each lexicon

Sentiment Score	Loughran	Bing	NRC
Mean	0.43	0.65	0.75
Standard Error	0.02	0.02	0.01
Standard Deviation	0.27	0.17	0.12
Sample Variance	0.07	0.03	0.01
Kurtosis	0.04	0.46	1.01
Skewness	-0.32	-0.53	-0.93
Range	1.33	0.98	0.59
Minimum	-0.33	0.02	0.38
Maximum	1.00	1.00	0.97
Sum	52.20	78.18	89.66
Count	120	120	120

Source: Own editing based on results

In *Table 2* we have the descriptive statistics for the article's effects sizes, meaning the actual results reported by the authors. The average mean for the effect sizes is 0.41, with a range between -0.10 and 1.35. For the effect sizes in theory the possible results would range from -1 to 1 as well, as most of the results are β (betas) of the studies, however there was one exception where the authors reported the β of a relationship to be 1.35. We decided to include that one in the study.

We see that the effect sizes on average are quite high. So at least from this finding we can state that on average big data analytics have a positive impact within the organization. However, to generalize this, there is a need for a more elaborated analysis which is called meta-analysis. This is beyond the purpose of the current study.

When we compare the descriptive of the two measures, sentiment scores and effect sizes we notice that the results match, at least on average. That's why the critical step for our study was the correlation analysis.

Table 2. Descriptive statistics for effect sizes per article based on each lexicon

Article's Effect sizes	
Mean	0.41
Standard Error	0.02
Standard Deviation	0.24
Sample Variance	0.06
Kurtosis	0.86
Skewness	0.70
Range	1.46
Minimum	-0.11
Maximum	1.35
Sum	49.80
Count	120

Source: Own editing based on results

5. RESULTS AND DISCUSSIONS

However, when doing the correlation analysis, the result is somehow surprising. As we can see from scatter plots in *Figure 3*, *Figure 4* and *Figure 5*, there is almost 0 correlation between the effect size reported by the authors and the sentiment score of the text based on the 3 lexicons Loughran, Bing and NRC.

The correlation coefficients between effect size and sentiment score depending on the lexicon used to score the sentiment are as it follows:

- Correlation coefficient between the effect size of the article and the sentiment score based on Loughran Lexicon: -0.091
- Correlation coefficient between the effect size of the article and the sentiment score based on Bing Lexicon: 0.092
- Correlation coefficient between the effect size of the article and the sentiment score based on NRC Lexicon: 0.087
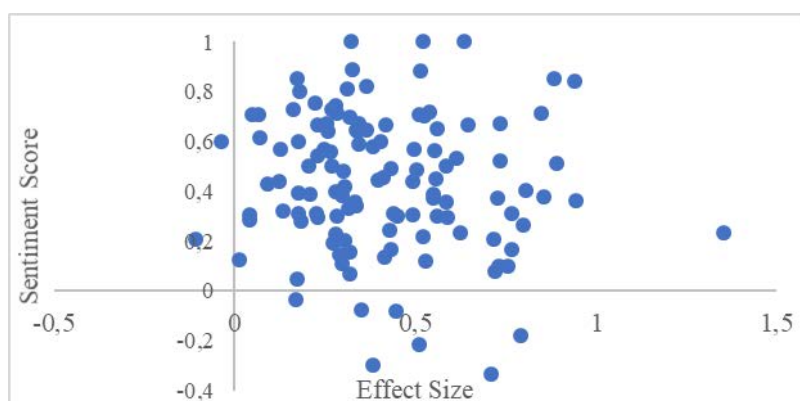

Figure 3. Effect size & Sentiment score: Loughran Lexicon
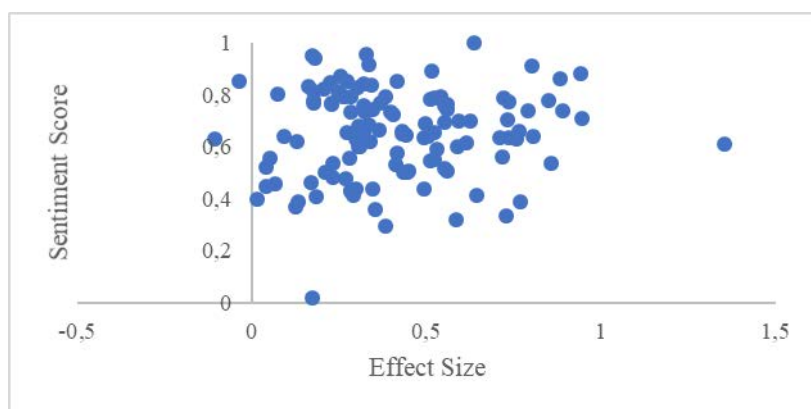*Source:* Own editing based on results

Figure 4. Effect size & Sentiment score: Bing Lexicon
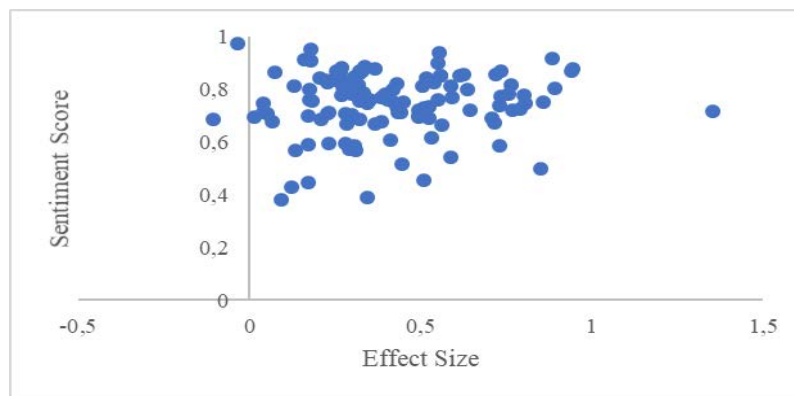*Source:* Own editing based on results

Figure 5. Effect size & Sentiment score: NRC Lexicon
Source: Own editing based on results

These correlation coefficients are very weak, whether positive or negative. To interpret this coefficient, we can say that the results demonstrate that there is no correlation at all between the two measures, the effect size reported by researchers and the sentiment score based on the text. This means that some of the articles that had an insignificant or mediocre result actually had a quite positive wording, while articles with stronger results had a more temperate or mild phrasing.

We think this happened because of a couple of reasons. First the hype of big data and data science in both the business world and the academia, and second because of the higher publishing rate of the articles with strong significant results, there might be a tendency among authors to present their studies in a brighter, more positive phrasing than the results themselves. We noticed that the average difference between the two scores, the sentiment score and the effect size is 0.196. This means that usually the sentiment score is higher than the effect size reported in studies.

When calculating the sentiment scores, because we used three lesions per each article this resulted in 360 observations (120 articles included in study). Out of this 269 had the sentiment score higher than the effect size for that article and only the remaining 91 recorded the opposite, effect size higher than sentiment score. This is an additional discovery which inclines us to believe that the results of the studies are usually overstated.

As a summary, we can say with no doubt, the initial hypothesis we made, that there is a positive correlation between the effect size of the study and the sentiment scoreis rejected. As we tested against the different lexicons, none of them seemed to result in a score similar with the effect size obtained within the study itself.

Although we did not have this initially in mind when we started the research, we believe these results are representative for this field of study, in the way that there is a disconnection between what is being said and what the results show. In the following paragraphs we would like to make the case for this judgement.

We assessed the sentiment of the articles using three lexicons, we detected that there is no correlation between the attitude of the authors and the results from their studies, at least not identifiable when analyzing using mathematical algorithms.

This raises further questions. First, could it be that the language from this research field tends to be to hyped? There are arguments to sustain this assumption, from both, the practitioner's world and the academia. In the practitioner's world there is a clear trend towards digitalization and automation. There is not one single industry, which we can think about, that is not currently being disrupted by all the digital advances. That's why some call this moment as the fourth industrial revolution. Data science and big data analytics are in the middle of this. We see operations, business models and strategies being influenced by these advances. Because of this, the pressure is huge for each business representative to pose as a big data ambassador. Otherwise it risks to create the impression that his company is lagging behind. In the academia, there is a similar pressure, as it

seems that the researcher's world is one step behind the business world in terms of theorizing and analyzing the impact of big data analytics. This creates the perfect climate for everybody to exaggerate the implications and the effects of big data analytics in the fear of looking outdated or not to understand this topic. But could it be that this is all deceptive and we will live a revelatory moment of the type of *"the emperor has no clothes"* from the Hans Christian Anderson's beloved story. Clearly, our study alone does not have the necessary extent to close the case one this question, rather to carry on the discussion. All we can pledge is that we see no connection between what researchers found (statistical results) and the way they express it (sentiment analysis).

Second, could it be that a sentiment analysis is not applicable for academic papers as they usually lack the emotional nuance? However, the range of the sentiment results exposed in both *Table1* and on the scatter plots from *Figure 3*, *Figure 4* and *Figure 5* confirm that we obtained a broad spectrum of results, especially on the positive side. This is further sustained by the standard deviation of the sentiment scores, 0.27, 0.17 and 0.12, depending on the lexicon used. So, there is actually multiple identifiable attitudes expressed within the articles we used. And especially because we used mainly the results part of the article, this gives us confidence that the scoring for each article should be significant for the attitude of the authors.

6. CONCLUSIONS

6.1 Contribution of the study

The aim of our study was to review the entire literature for the studies which tested the relationship between big data analytics and firm performance. In order to get as many articles as possible, we used Publish or Perish, which is a "search engine" that uses Google Scholar's database. Next, we filtered only the articles that met our criteria, so in the end we tested our hypothesis on a sample of 120 articles. By reviewing the sample of articles, we identify that on average the relationship between a form of big data analytics and firm performance is positive. However, in order to confirm this, further analysis is recommended and this was not the purpose of the study.

As for our main purpose, we wanted to test whether a sentiment analysis on the articles included in the sample, would result in a score per each article, which will correlate with the effect sizes reported by the authors. The sentiment score was calculated as a ratio between the positive and the negative words used in the article. In other words, we expected that the studies with higher and stronger results would have a more positive phrasing.

When testing the hypothesis, we found absolutely no relationship between the two elements: sentiment score and effect size of the study. In order to obtain the sentimentscore, we tested against three lexicons: Loughran, Bing and NRC. The correlation coefficients are as it follows: (1) effect size vs. sentiment score based on Loughran lexicon -0.091; (2) effect size vs. sentiment score based on BING lexicon 0.092; (3) effect size vs. sentiment score based on NRC lexicon 0.087.

As the main finding we can declare that there is no relationship between the effect size of the article and the sentiment score (determined based on the wording). Additionally, we identified that the number of the articles which have a more positive phrasing than the results themselves is considerably higher (269 vs. 91). So about 75% of the articles have an overstated language compared to the results. Also, we found the there are articles where the language is more *"moderate"* although they have strong results. So not everybody *"amplified"* their results. However, on average the extent to which the sentiment score for the *"amplified"* ones differs from the effect size is higher (0.260) vs. the extent to which the sentiment score for the *"moderated"* ones differs from the effect size (-0.064).

As a summary, we found based on the literature review and the results from the articles we included in our analysis, that there is an alignment when it comes to the question of whether big data analytics positively impacts firm performance. However, the methods and the perspectives upon what big data analytics and firm performance mean are divided. And more important we found no

connection between the results themselves from each article and the way in which these findings are presented in terms of phrasing. The tendency is to overstate the results.

6.2 Limitations and future research

We would like to think that our research is somewhere at the limit between qualitative and quantitative research. In the way that we aimed to review the text of the literature, in our case the sample of 120 articles. However, we did it in a more objective, impartial way, by using an algorithm which determined the sentiment score per each article and on the other side we extracted the effect sizes of each article, meaning the results reported by the authors. This allowed us to test our hypothesis.

This allowed us to test only the relationship between sentiment scores and effect sizes. So, we cannot estimate the impact of big data analytics upon firm performance. In order to make solid statements about this, a meta-analysis study would be suitable. This type of study is usually used in order to settle a debate within a field of study.

Another limitation is due to the possibility of missing relevant studies. We tried our best to include all the studies, by using an article *"search engine"* on Google Scholar database, however, some of the articles we needed, might have never been published. This is due to publication bias, and is explained by the fact that studies with strong and positive results have a higher success rate to be published.

Third, we tested our hypotheses on the overall sample at once. By creating sub-groups based on elements such as type of big data analytics, type of performance, industry or firm size we might obtain different results per each sub-group.

REFERENCES

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R. & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment?. *International Journal of Production Economics*, *182*, 113-131.

Aydiner, A. S., Tatoglu, E., Bayraktar, E., Zaim, S. & Delen, D. (2019). Business analytics and firm performance: The mediating role of business process performance. *Journal of Business Research*, *96*, 228-237.

Bag, S., Wood, L. C., Xu, L., Dhamija, P. & Kayikci, Y. (2020). Big data analytics as an operational excellence approach to enhance sustainable supply chain performance. *Resources, Conservation and Recycling*, *153*, 104559.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120.

Cao, G., Duan, Y. & Li, G. (2015). Linking business analytics to decision making effectiveness: A path model analysis. *IEEE Transactions on Engineering Management*, *62*(3), 384-395.

Côrte-Real, N., Oliveira, T. & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, *70*, 379-390.

Duan, Y., Cao, G. & Edwards, J. S. (2020). Understanding the impact of business analytics on innovation. *European Journal of Operational Research*, *281*(3), 673-686.

Dubey, R., Gunasekaran, A., Childe, S. J., Fosso Wamba, S., Roubaud, D. & Foropon, C. (2019). Empirical investigation of data analytics capability and organizational flexibility as complements to supply chain resilience. *International Journal of Production Research*, 1-19.

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F. et al. (2019). Can big data and predictive analytics improve social and environmental sustainability?. *Technological Forecasting and Social Change*, *144*, 534-545.

Ghasemaghaei, M. (2018). Improving organizational performance through the use of big data. *Journal of Computer Information Systems*, 1-14.

Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency. *Decision Support Systems*, *120*, 14-24.

Ghasemaghaei, M., Hassanein, K. & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95-105.

Gupta, M. & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049-1064.

Gupta, S., Qian, X., Bhushan, B. & Luo, Z. (2019). Role of cloud ERP and big data on firm performance: a dynamic capability view theory perspective. *Management Decision*.

Hao, S., Zhang, H. & Song, M. (2019). Big data, big data analytics capability, and sustainable innovation performance. *Sustainability*, *11*(24), 7145.

Hosoya, R. & Kamioka, T. (2018). Understanding how the ad hoc use of big data analytics impacts agility: a sensemaking-based model. *Paper presented at 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*. Durban, South Africa.

Huang, C. K., Wang, T. & Huang, T. Y. (2020). Initial evidence on the impact of big data implementation on firm performance. *Information Systems Frontiers*, *22*(2), 475-487.

Jeble, S., Dubey, R., Childe, S. J., Papadopoulos, T., Roubaud, D. & Prakash, A. (2018). Impact of big data and predictive analytics capability on supply chain sustainability. *The International Journal of Logistics Management*, *29*(2), 513-538.

Kwartler, T. (2017). *Text mining in practice with R*. Chichester, UK: John Wiley & Sons.

Loughran, T. & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J. & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60-68.

Niebel, T., Rasel, F. & Viete, S. (2019). BIG data–BIG gains? Understanding the link between big data analytics and innovation. *Economics of Innovation and New Technology*, *28*(3), 296-316.

Popovič, A., Puklavec, B. & Oliveira, T. (2019). Justifying business intelligence systems adoption in SMEs. *Industrial Management & Data Systems*, *119*(1), 210-228.

Raguseo, E. & Vitari, C. (2018). Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research*, *56*(15), 5206-5221.

Silge, J. & Robinson, D. (2017). *Text mining with R: A tidy approach*. Sebastopol, CA: O'Reilly Media, Inc.

Singh, N. P., & Singh, S. (2019). Building supply chain risk resilience. *Benchmarking: An International Journal*, *26*(7), 2318-2342.

Tang, C. P., Huang, T. C. K. & Wang, S. T. (2018). The impact of Internet of things implementation on firm performance. *Telematics and Informatics*, *35*(7), 2038-2053.

Thirathon, U., Wieder, B., Matolcsy, Z. & Ossimitz, M. L. (2017). Impact of big data analytics on decision making and performance. *Paper presented at International Conference on Enterprise Systems, Accounting and Logistics*. Porto, Portugal.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R. & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365.

Yu, W., Chavez, R., Jacobs, M. A. & Feng, M. (2018). Data-driven supply chain capabilities and performance: A resource-based view. *Transportation Research Part E: logistics and transportation review*, *114*, 371-385.