Unfolding the relations between companies and technologies under the Big Data umbrella

Abstract

Big Data is dominating the landscape as data originated in many sources keeps piling up. Information Technology (IT) business companies are making tremendous efforts to keep the pace with this wave of innovative technologies. This study aims to identify how the different IT companies are aligned with emerging Big Data technologies. The approach consisted in analyzing 11,505 news published between 2013 and 2016 and aggregated through Google News. The companies were categorized according to their position in the 2017 Gartner Magic Quadrant for advanced analytics. A text mining and topic modeling procedure assisted in summarizing the main findings. Leaders dominated a large fraction of the published news. Challengers are making a significant effort in investing in predictive analytics, overlooking other technologies such as those related to data preparation and integration. The results helped to shed light on the emerging field of Big Data from a corporate perspective.

Keywords

Big Data companies; Big Data technologies; online news; Gartner Magic Quadrant.

1. Introduction

The development of disruptive technologies such as social networking, cloud computing, and the Internet-of-Things led to a continuous increase of data and its accumulation at an incalculable speed (Lapalme et al., 2016). The abovementioned factors contributed to the trivialization of a new concept: Big Data (Meng, 2013). This area of research has emerged in a diverse spectrum of technological innovations and opportunities made available by the information revolution (Romero & Vernadat, 2016). The expectations that Big Data will lead today's society to a new and captivating age of innovation are high (Goes, 2014).

There are several techniques currently associated with the term Big Data, some even older than the appearance of the theme itself (Sharda et al., 2018). This proliferation makes it hard to get a clear picture of how companies are following Big Data approaches. This study aims to address such challenge by understanding what are the main Big Data techniques embraced by each type of company. As a source of information, this study uses online news recently published. Specifically, the Google News aggregator was chosen, through which it was possible to gather via web scraping a total of 11,505 news (from 2013 to 2016). Thus, this study provides a business perspective of a highly addressed topic of research. Such perspective enables to highlight the needs, trends and gaps of companies under the Big Data realm, helping in guiding future applied research under a business context.

The proposed approach was framed under two dimensions: (1) technological companies offering Big Data solutions, and (2) Big Data technologies. Selecting the most meaningful technologies as well as the most relevant companies that address such a vibrant domain is a subjective task limiting the scope; however, it is essential for the feasibility of the proposed analysis. To reduce such subjectivity, the Gartner Magic Quadrant for Advanced Analytics was used to identify the most relevant companies. The Magic Quadrants published by Gartner are often adopted by scholars (e.g., Bāliņa et al., 2016; Kretzer et al., 2014) and encompass a group of companies with strong relevance in world economy. As an example, Graves et al. (2015) adopted the Gartner Quadrant on cloud storage to choose a grounded sample of challengers and leaders for testing their cloud services. The fact that the Gartner Magic Quadrant already aggregates companies into four groups (Leaders, Visionaries, Niche Players, and Challengers) provides a focused categorization in four quadrants of the companies offering services in advanced analytics.

The vast number of technologies needed to be narrowed to the most significant in Big Data. Such selection was grounded on a study published by the Forrester's TechRadar (Yuhanna & Hopkins, 2016) where the authors evaluated the maturity and evolution of 22 Big Data technologies. Our study adopted a text mining-based approach considering seven of those technologies which produce a medium or high business value, plus those with a predictable future success trajectory and finally those that were in the survival and growth stage, namely: predictive analytics, nosql databases, stream analytics, data virtualization, data integration, data preparation, and data quality.

2. Background

2.1.Big Data

Over the years, the term Big Data has increasingly been used in scientific papers (Figure 1). After 2011, there was an exponential growth in the written use of the expression “Big Data”. Accordingly, this concept is currently used daily, as it is an expression commonly trivialized by the mass media, although demanding a deeper description to contextualize its use.

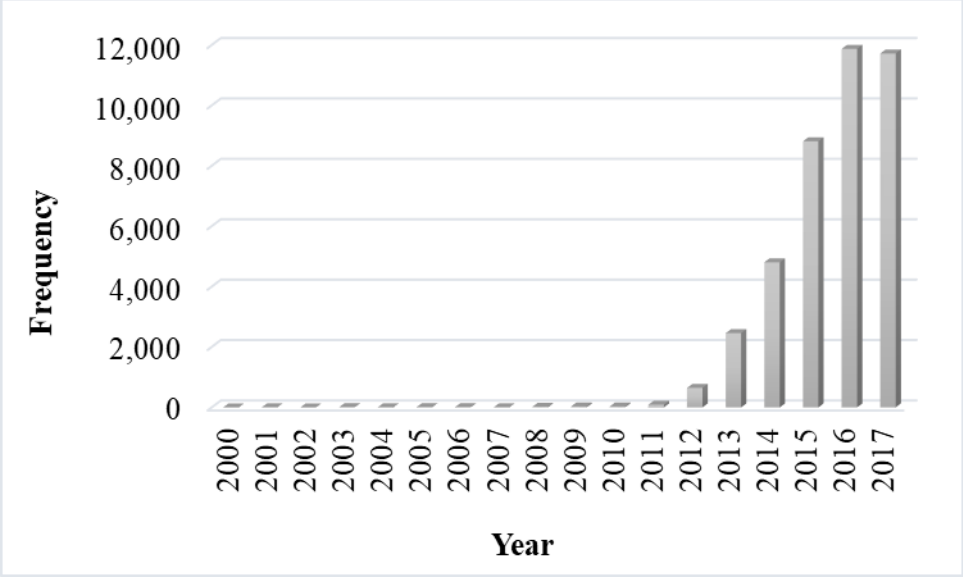

Figure 1 - Frequency of documents with the term Big Data in Scopus.

According to Hashem et al. (2015, p. 9), “Big data is a term utilized to refer to the increase in the volume of data that are difficult to store, process, and analyze through traditional

database technologies". Laney (2001) suggested that Volume, Variety, and Velocity (or the three V's) were the three dimensions of data management challenges. Hence, the three V's emerged as a common framework for describing Big Data (Chen et al., 2012). Indeed, Big Data represents large-volume, high-speed, and high-variety data that requires efficient and innovative forms of information processing to extract decision support and process automation content (Gartner IT Glossary, 2017).

Big Data volume (data magnitude) definitions are relative, depending on factors such as time and data type. What can be considered Big Data today may not be in the future, as data storage capacities will increase, which will allow the collection of larger datasets. Variety, another Big Data property, refers to the heterogeneous structure of a dataset. Technological advances allow companies to use various types of data, whether structured (e.g., relational tables), semi-structured (e.g., XML) or unstructured (e.g., text, image, audio, and video). Velocity refers to the rate that the data is generated and the speed that it should be analyzed and put into practice. Recently, the number of V's increased with the need to better frame Big Data. Accordingly, Seddon and Currie (2017) have developed a model with four additional V's: variability, veracity, visualization, and value. The abovementioned characteristics leveraged existing approaches (e.g., predictive analytics) or gave rise to novel ones (NoSQL databases) to handling data. Thus, Big Data is an umbrella term covering all of them under a new philosophy devoted to dealing with this phenomenon (Chen et al., 2012).

Nowadays, companies are flooded with data from both internal and external sources. To thrive in todays' competitive environment, they need to fully apprehend insightful knowledge able to leverage their businesses. Therefore, most of large corporations are embracing Big Data and taking the most from it in domains such as marketing (Amado et al., 2017) and finance (Fang & Zhang, 2016).

2.2.Gartner Magic Quadrant

Gartner introduced Gartner's Magic Quadrant (GMQ) research methodology to help understand technology providers. The methodology tries to support investment opportunity by answering the following question: which are the competing players in the major technology markets and how are they positioned? Magic Quadrants provide a

graphical summary of the maturity and direction of technology providers in markets where growth is high and provider differentiation is distinct (Black & Thomas, 2013).

Vendors are compared based essentially on two Gartner's criteria: completeness of vision and ability to execute. Gartner's two criteria cover a large set of topics such as (Black & Thomas, 2013): Market Understanding, Marketing Strategy, Product/Service, Overall Viability, etc.

In a Magic Quadrant, a graphical positioning is provided considering four types of technology suppliers (Black & Thomas, 2013; Figure 2 from source):

- "Leaders execute well against their current vision and are well positioned for tomorrow;
- Visionaries understand where the market is going or have a vision for changing market rules, but do not yet execute well;
- Niche Players focus successfully on a small segment, or are unfocused and do not out-innovate or outperform others;
- Challengers execute well today or may dominate a large segment, but do not demonstrate an understanding of market direction."

Every year Gartner releases Magic Quadrant reports for specific technologies. For example, in 2017 Gartner released a Magic Quadrant for Advanced Analytics report (Figure 2). The report focuses on the evaluation of providers of advanced analytics platforms that use them to build extensive solutions. Although Gartner does not publish yet a specific Big Data Quadrant, both subjects (Big Data and advanced analytics) are tightly coupled as companies seek to take advantage of Big Data sources adopting data-driven analytical approaches (Barton & Court, 2012). Wal-Mart is an example of a large company taking advantage of data analytics to exploit the knowledge hidden in purchase records from their point-of-sale terminals (summing up to 267 million transactions per day) to improve pricing strategies and advertising campaigns (Chen & Zhang, 2014).

Figure 2 - GMQ for Advanced Analytics Platforms (source: Gartner Magic Quadrant for Advanced Analytics, 2017).

Gartner Magic Quadrants have been extensively adopted for research purposes. Kretzer al. (2014) used the 2014 Magic Quadrant for Business Intelligence and Analytics (BI&A) platforms report to select companies for analyzing which factors prevented stable BI&A platforms from enabling organizational agility. Chen et al. (2012) published an influential article supported on the Magic Quadrant for BI&A providing a framework that identified the evolution, applications, and emerging research areas of BI&A, in terms of their key characteristics and capabilities.

2.3.Text Analytics and Text Mining

A significant portion of unstructured content collected by an organization is in textual format, from e-mail communications and corporate documents to web pages and social media content (Ittoo et al., 2016). Text Analytics and Text Mining technologies aim to extract information from textual data and using it for research or business purposes. Both technologies share the same methods and tools, and their difference is more related with

the background of the professionals who use them. Nevertheless, according to Gartner (Gartner IT Glossary, 2017), Text Mining focuses on the process of extracting information from textual data collections while Text Analytics refers to the process of removing information from text sources.

Text Analytics has its roots in information retrieval and computational linguistics (Chen et al., 2012). In information retrieval, document representation and query processing are the groundwork for the development of the vector space model, Boolean recovery model and probabilistic recovery model, which has become the basis for modern library search tools and enterprise search systems (Salton, 1997).

News feeds from a social network, emails, blogs, online forums, questionnaire response and news are some examples of textual data. Text mining (similar to Text Analytics, but more heuristic-driven, focused on the examination of the structure, and less algorithmic) involves statistical analysis, computational linguistics and machine learning (Gandomi & Haider, 2015).

Text Mining (TM) tools allow organizations to convert large volumes of human-generated text into a simple summary using qualitative textual analysis tools where the number of occurrences of a text with certain relevance is counted, which after applying quantitative methods to extract knowledge, serves as an instrument in evidence-based decision-making. For instance, one of its applications is in forecasting the stock market, based on information extracted from financial news (Oliveira et al., 2016).

3. Materials and methods

3.1.Data collection

For the collection of news, Google News was chosen to start from a single site with all the news about a specific topic. It was defined that the collected news would be between 2013 and 2016 in the English language, and contain the expression "Big Data" in the title. Google News benefits from Google's extensive expertise in websites' prioritization, meaning that the most relevant news appear first (sorted by relevance). To obtain a monthly representative subset of news, twelve queries (one per month) were executed.

The news extraction was conducted in February 2017 and it consisted of two stages: Partial News Extraction and Python Script for Extraction.

From each news, the following fields were extracted: URL Google News link, URL source page, extraction date, title, source, date, and summary (Figure 3).

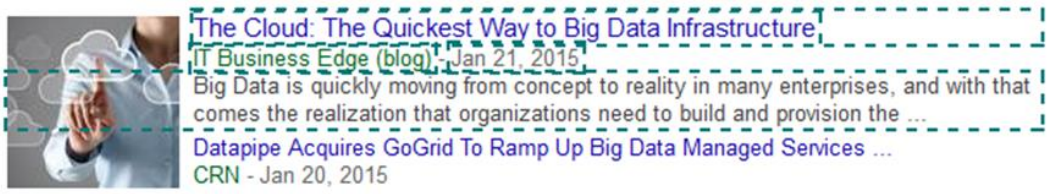

Figure 3 - Fields extracted (title, source, date, and summary).

Based on the URL for each news source page, the procedure required diving into each website of all those different sources where the news was effectively published to extract all text from it. Some of the news could not be obtained due to restrictions in source pages. Thus, all news from which the obtained text contained less than 100 characters were discarded, as those typically contained error or forbidden access messages. The result of this process was a comma-separated file containing 11,505 rows, one per news. Table 1 shows the top 20 Big Data news sources.

Table 1 - Top 20 Big Data news sources.

Source	Source News Frequency	% Source News
Forbes	678	5.89%
TechTarget	249	2.16%
Smart Data Collective	230	2.00%
VentureBeat	220	1.91%
PR Newswire (press release)	197	1.71%
InformationWeek	179	1.56%
Health IT Analytics	162	1.41%
TechRepublic	146	1.27%
Business Wire (press release)	134	1.16%
ZDNet	128	1.11%
insideBIGDATA	124	1.08%
Information Age	111	0.96%
Datanami	108	0.94%
TechCrunch	97	0.84%
Huffington Post	94	0.82%
V3.co.uk	94	0.82%
Wall Street Journal (blog)	92	0.80%
DZone News	91	0.79%
ITProPortal	86	0.75%
CIO	83	0.72%

3.2.News' analysis approach and dictionary definition

The undertaken analysis focused mainly on extracting hidden patterns of textual knowledge from the news dataset using Big Data Technologies (BDTS) and Gartner's Magic Quadrant (GMQ) as the two categorizing dimensions (Figure 4). It was crucial to feed the extraction procedure with a lexicon dictionary established with relevant terms to the definition of the BDTS and the GMQ. Thus, the process of data cleaning and conversion used the lexicon contained in this dictionary to reduce the news text in sets of relevant terms. The lexicon dictionary from both BDTS and GMQ terms was compiled as a single input, to make it possible for the model to cross both domains. The main text mining procedure output is the document-term matrix. This matrix has two dimensions: Big Data news and each of the terms considered. Each cell contains the frequency in which each term occurs in each news. A preliminary analysis was conducted using a frequency table, where the number of term occurrences was counted, and a word cloud, which presents an easier way of understanding these occurrences.

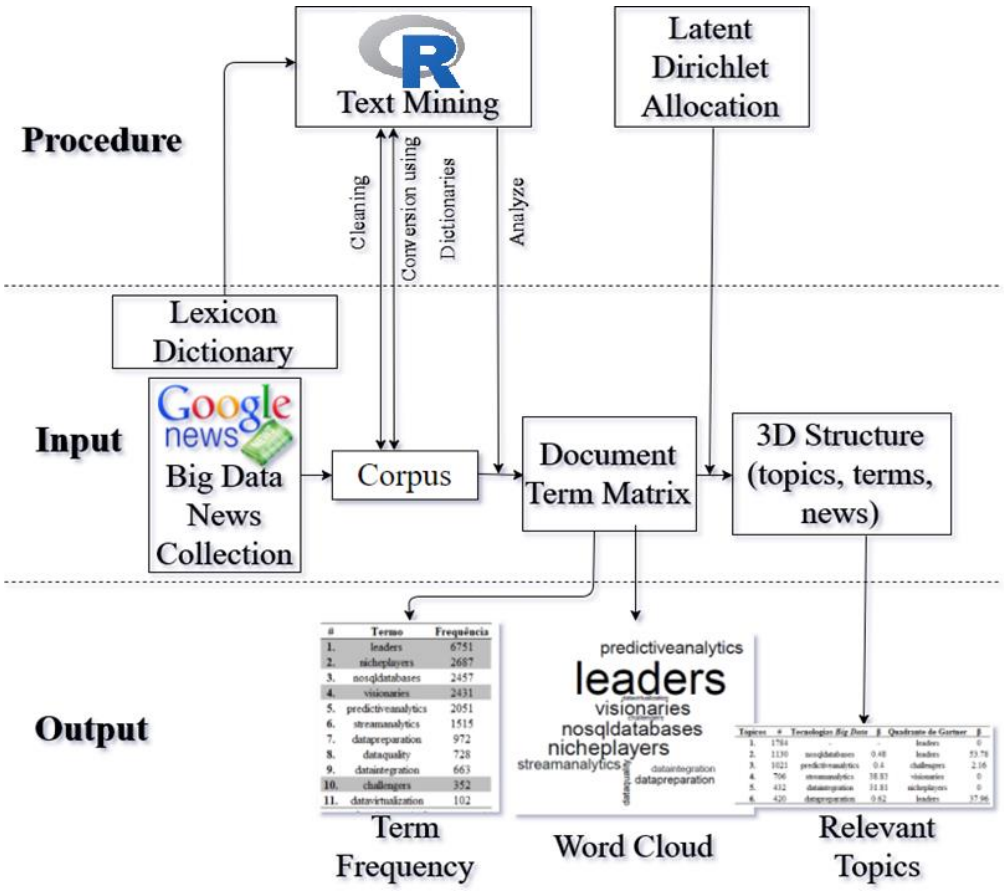

Figure 4 - Undertaken approach.

To obtain the most relevant topics, the document-term matrix serves as input to the latent Dirichlet allocation (LDA) topic modeling (Blei et al., 2003). The final output of the LDA is a three-dimensional structure that comprehends terms, news, and topics. Therefore, it is possible to obtain for each topic a relation with the terms of the dictionary, through a distribution β. It is also possible to observe, for each news, which topic better expresses it. As it was previously explained, two distinct dictionaries were implemented, one in the BDTS domain based (Table 2) and the other composed by the Big Data companies (Table 3). For search purposes only, all values appear in lower case, including the names of the companies.

Table 2 - BDTS dictionary (partial)

Reduced Term	Similar or Domain Term
predictive analytics	abm,actian analytics platform,advanced miner,...
nosql databases	nosql,hadoop/hbase,cassandra,hypertable,...
stream analytics	apache flink,spark streaming,...
data virtualization	actifio sky,datacurrent,denodo,...
data integration	actian dataconnect,analyza,...
data preparation	platfora,paxata,datawatch, tamr platform,...
data quality	acquiro leadmatch,clear analytics,...

Table 3 - GMQ dictionary.

Reduced Term	Similar or Domain Term
leaders	ibm,sas,rapidminer,knime
visionaries	microsoft,h2o ai,dataiku,domino data lab,alpine data
nicheplayers	fico,sap,teradata

challengers	mathworks,quest,alteryx,angoss

The technologies that compose BDTS from Table 2 were chosen based on their business added value adjusted for uncertainty (based on its potential impact, feedback and evidence of implementation and market reputation), future trajectory and ecosystem phase. Particularly, their business and life cycle value was based on the report published by Yuhanna and Hopkins (2016). Only technologies with a business added value adjusted for uncertainty with a medium or high value, technologies with a future trajectory with significant success, represented by the upper curve in Figure 5 and the technologies in a stage of survival or growth were considered. From the resulting options, the BDTS that were included in the BDTS dictionary are highlighted with a surrounding rectangle.

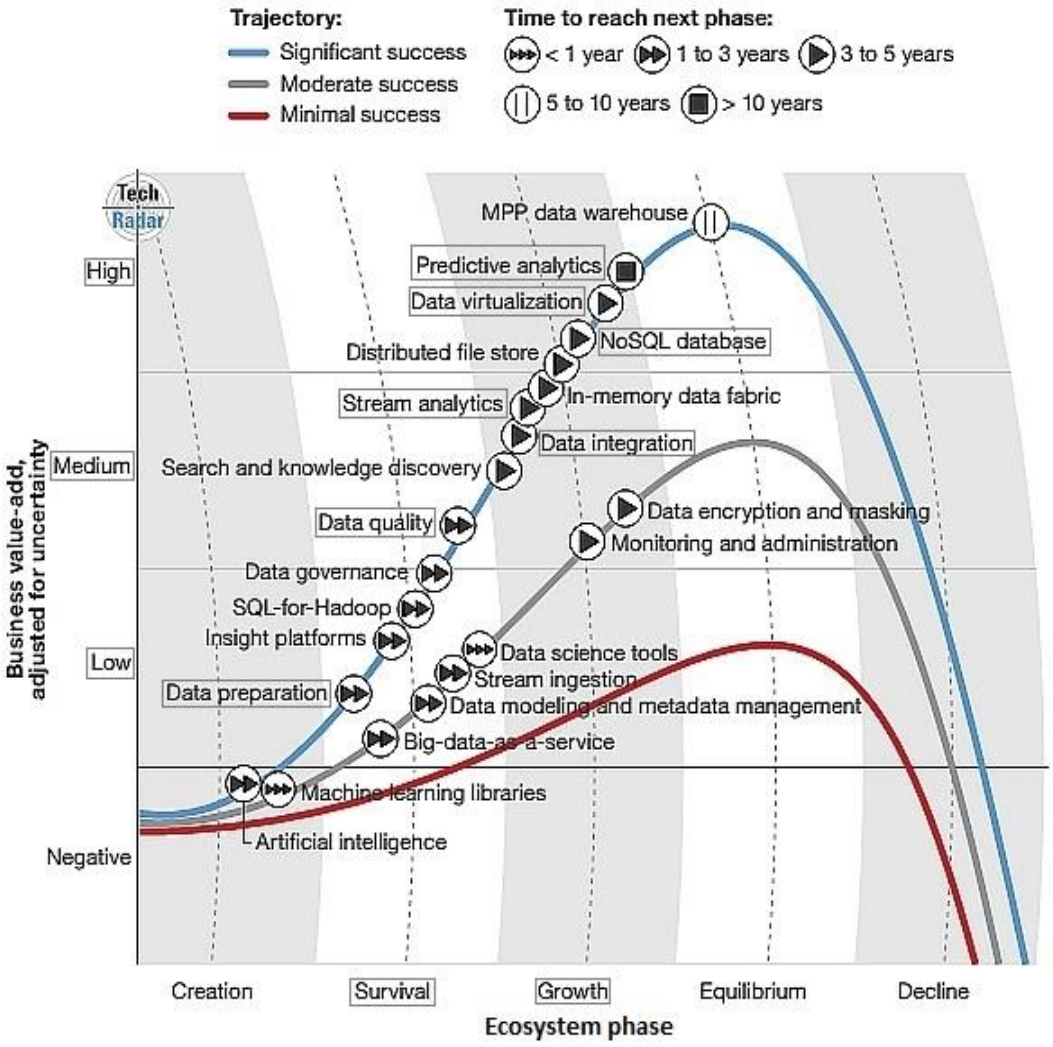

Figure 5 - BDTS according to their value to the Business and Life Cycle (source: Yuhanna & Hopkins, 2016).

The similar or domain terms observed in Table 2 were chosen from business independent reviews' sites. This provided coherent lexica known and adopted under companies' context without the bias of using terms directly obtained from companies' information sources (e.g., institutional websites). The chosen sites were:

- Predictive analytics http://www.predictiveanalyticstoday.com/top-predictive-analytics-software/
- NoSQL databases http://bigdata-madesimple.com/a-deep-dive-into-nosql-a-complete-list-of-nosql-databases/
- Stream analytics http://www.predictiveanalyticstoday.com/top-open-source-commercial-stream-analytics-platforms/
- Data virtualization https://www.trustradius.com/data-virtualization
- Data integration https://www.trustradius.com/data-integration?f=50&o=alpha&s=25
- Data preparation http://www.predictiveanalyticstoday.com/data-preparation-tools-and-platforms/
- Data quality https://www.trustradius.com/data-quality?o=alpha

GMQ was used to define the dictionary of Big Data companies, which results from a specific market research, providing a broad graphical view of the competitors' positions in that market, as explained in Section 2.2.

3.3.Topic modeling

To perform the TM procedure, the statistical open-source tool R was adopted. Namely, the "tm" and "topicmodel" packages were chosen. The former provides text mining functions, while the latter implements the LDA algorithm (Calheiros et al., 2017). The LDA algorithm is a three-level hierarchical Bayesian modeling process, which groups a set of items into topics defined by words or terms, where each of the terms identified characterize a topic (Blei, 2012). LDA allows to analyze the relative relevance of each term using the distribution β value, which characterizes the relationship between the topic and the specified term. A β close to zero represents a stronger relationship between the term and its corresponding topic. Since both dictionaries are merged, this may mean that a topic can be best characterized by one of the terms related to a single category. However,

this technique also provides interesting insights into the relations between categories (Moro and Rita, 2018). This algorithm can be implemented and computed with only two parameters, the number of topics and the document-term matrix created by the TM procedure. The "ldatuning" package was used to identify the ideal number of topics, between four to eight. With experiments being made among the various possibilities, six was the perfect choice for the number of topics.

4. Results and discussion

Table 4 shows the number of occurrences of each term according to the dictionaries drawn on Table 2 and Table 3. The terms highlighted in gray represent Gartner Quadrants. The news extracted emphasize leaders' efforts in consolidating their market position through this type of mass media, appearing 6,751 times in 11,505 news. Also, it is possible to confirm that most news were focused on companies, as three of the four GMQ emerge in the first positions. Thus, news adopted a more corporate perspective rather than a technical one. Although the presence of some business media sources (e.g., Forbes) could explain such result, Table 1 also shows a large number of technical sources.

Table 4 - Terms' frequency.

#	Term	Frequencies
1.	leaders	6,751
2.	niche players	2,687
3.	nosql databases	2,457
4.	visionaries	2,431
5.	predictive analytics	2,051
6.	stream analytics	1,515
7.	data preparation	972
8.	data quality	728
9.	data integration	663
10.	challengers	352
11.	data virtualization	102

Note: The GMQ terms are identified in gray for better identification.

Figure 6 shows a visual representation of the highlighted results in Table 4 through a word cloud. The results exhibited account for 11 terms in total, demonstrating that the leaders are unequivocally the most prominent, according to the news scrupulously

scrutinized. From the technology perspective, nosql databases and predictive analytics are taking the most attention, although closely followed by stream analytics.

Figure 6 - Word cloud.

The topics unveiled through modeling are displayed in Table 5. Each topic is represented by a horizontal line, with the column marked as # counting the news that best matched the corresponding topic. The table also shows the most relevant BDTS and its GMQ counterpart that best characterize each topic. The β distribution value is a measure of how close the term is from the corresponding topic. Thus, a value closer to zero represents a stronger relation to that topic. Figure 7 shows the topics rearranged over a map centered on the four GMQ and the technologies orbiting them.

Table 5 - Topics discovered.

Topics	#	Big Data technologies	β	Gartner Quadrant	β
1.	1,784	-	-	leaders	0.00
2.	1,130	nosql databases	0.48	leaders	53.78
3.	1,021	predictive analytics	0.40	challengers	2.16
4.	706	stream analytics	38.83	visionaries	0.00
5.	432	data integration	31.81	niche players	0.00
6.	420	data preparation	0.62	leaders	37.96

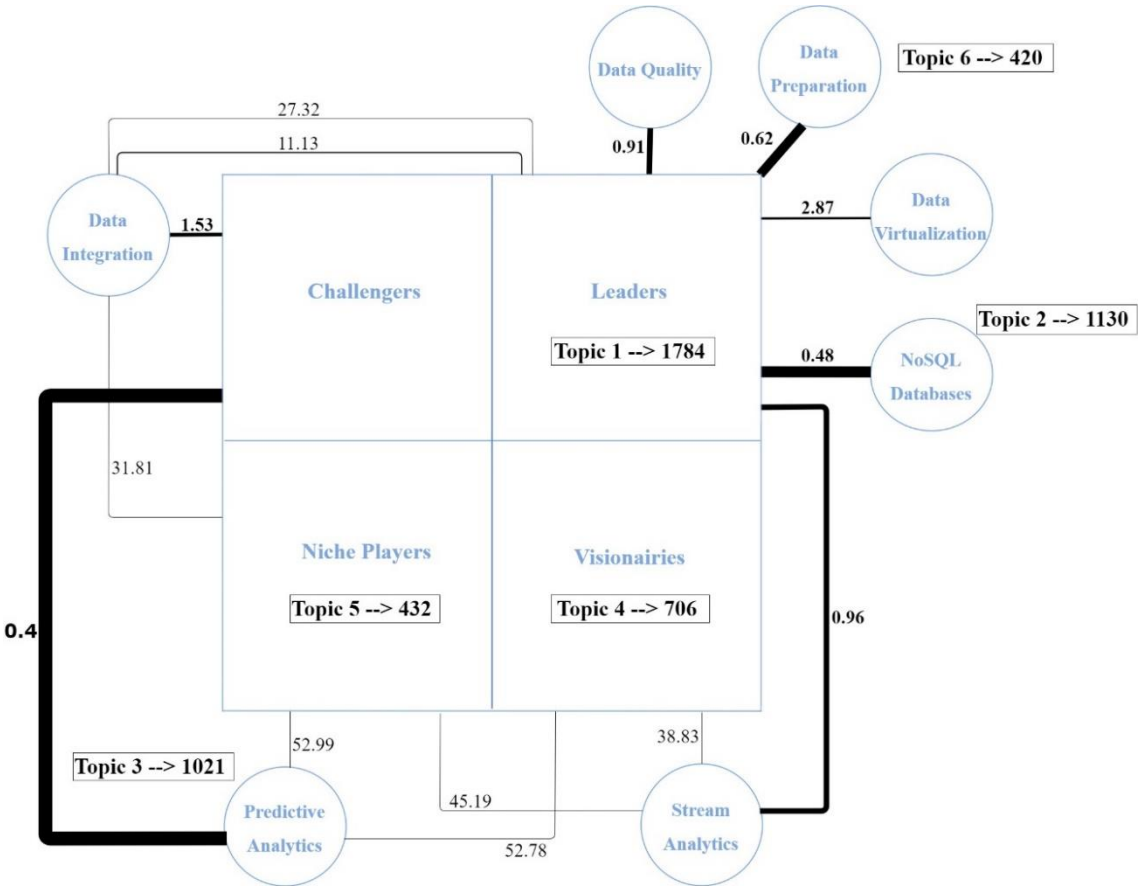

Figure 7 - Topics map.

The first striking discovery is the over-representativity of the leader companies, who dominate the landscape across the analyzed news. While this is a result of their high frequency (see Table 4), it also shows that these companies are spread throughout the news, as their corresponding quadrant is the most significant to three of the six discovered topics. Also, leaders have the "size advantage", since these are large enough to be able to invest simultaneously in several technologies, therefore emerging in three topics. Nevertheless, the matching is only perfect for the first topic, whereas for topics two and six the relation is tenuous, especially if compared to the relations found in the studies by Moro et al. (2017) and Moro and Rita (2018). This observation as well as the β values shown for all topics, immediately imply that there are two types of news: those specifically focused on the corporate perspective (topics 1, 4, and 5) and those adopting a technological perspective (topics 2 and 6). The interesting exception is topic 3 which reflects challengers betting on predictive analytics. According to Gartner, challengers reflect current trends but do not demonstrate an understanding of the market direction. Thus, this finding suggests there will be other Big Data technologies requiring further attention in future besides predictive analytics. As volumes of data keep increasing in

result of social media and the Internet-of-Things, there will be even more need to integrate and prepare such data to be useful for extracting valuable insights (Janssen et al., 2017). In fact, both niche players and leaders seem to have already understood such market forthcoming need, and news about data integration (topic 5) and data preparation (topic 6) are more related to these companies than to the remaining ones. However, long-term prospects can still reserve a significant role for predictive analytics as the world becomes less hunger for data and more hunger for valuable decision support (Chen et al., 2012). It would be interesting to repeat the undertaken empirical exercise in the future using forthcoming GMQ to see how companies shift their efforts to meet world needs for handling Big Data. Finally, visionaries are particularly investing on a novel trend called stream analytics. This refers to analyzing in real-time multiple streams of data from sources where the data is reaching at such a high velocity and volume that it needs to be promptly analyzed since it cannot be stored for future processing (Bifet & Réseaux, 2015).

Figure 8 shows that scholars have still devoted little attention to both stream analytics and data virtualization. However, these are emerging trends (Figure 6) to which leaders are already turning their attention (e.g., Salazar, 2014). Thus, it looks that there is plenty of road to go for researchers to meet the industry, as leaders set pace in these domains. Although the interest in data preparation has increased in recent years, it is not likely meeting industry leaders' requirements, as these are highly investing on related technologies. This appears as an interesting area where research gaps still exist and can be further explored with a proper alignment between scholars and the industry. Data quality and data integration are mature technological areas starting back before the Big Data hype. Research seems to have reached a plateau in data integration, which is aligned with the relative less attention that every company in the four Gartner Quadrants has devoted to it when compared to other BDTS (Figure 7). However, data quality, which is tightly associated with leaders, is receiving increased attention from scholars in the most recent years, showing an alignment with industry needs.

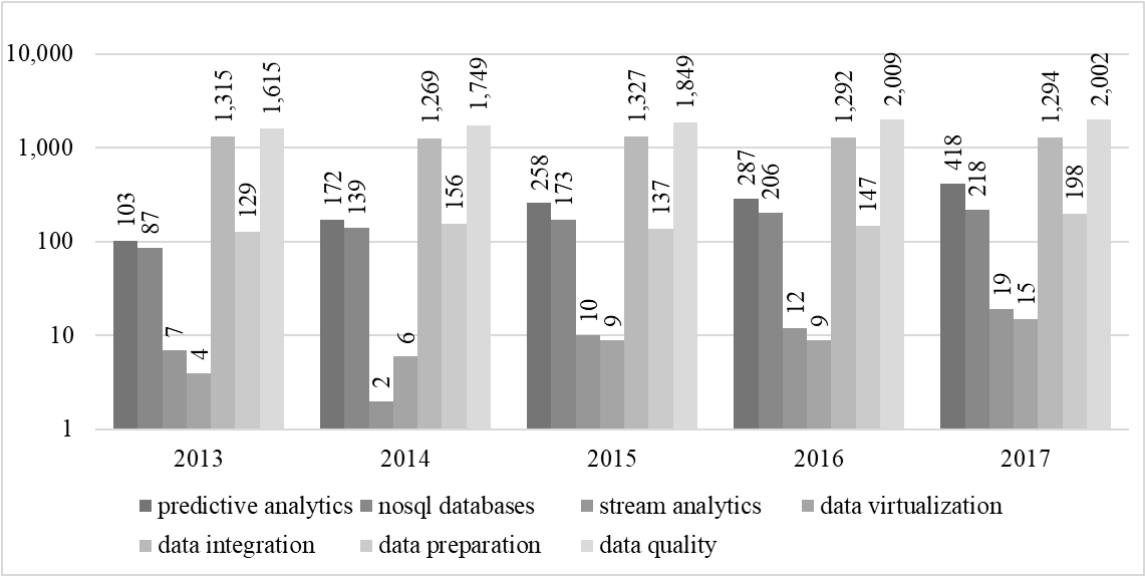

Figure 8 – Prevalence of Big Data Technologies in Scopus indexed documents.

5. Conclusions

This study provided interesting insights such as: (1) the leading companies do not focus on a single technology, but on the integration of several of them; (2) challengers are strongly investing in predictive analytics, although neglecting other more promising technologies and trends in terms of market demand for the near future (e.g., data preparation and integration).

The trends are constantly shifting, but news recently published pointed to the relevance of data preparation and integration, nosql databases, and predictive analytics. Nevertheless, challenging companies seem to be investing more on the latter, which, according to Garner's definition of challengers, indicates that the market is currently pointing to different directions. Also, stream analytics seems to be an emerging trend with visionaries investing in it, suggesting that there may appear promising related technologies in the future.

Big Data companies and technologies are evolving very quickly. In addition, some news may be a result of companies pushing hard to get their products published, and this effect may be amplified for large leading companies. Although the Gartner Magic Quadrants are well-known and widely disseminated rankings, it would be interesting to compare the achieved results with other rankings, in order to assess the consistency of Gartner's categorization in identifying the main market trends in the vibrant theme of Big Data.

This study calls for more research projects with coupled partnerships between the industry and research centers on Big Data technologies. A proper alignment between industry needs and scholars is likely to lead to an increase in innovation in recent technologies such as stream analytics, translated into an expected number of related articles published in the near future. Additionally, surveys may be applied to Big Data professionals to assess if their opinions are convergent with the conclusions drawn from this study.

Competing interests statement

The authors have no competing interests to declare.

References

Amado, A., Cortez, P., Rita, P., & Moro, S. (2017). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. European Research on Management and Business Economics, 24(1),1-7.

Bāliņa, S., Žuka, R., & Krasts, J. (2016). Opportunities for the Use of Business Data Analysis Technologies. Economics and Business, 28(1),20-25.

Barton, D., & Court, D. (2012). Making advanced analytics work for you. Harvard Business Review, 90(10),78-83.

Bifet, A., & Réseaux, D. I. (2015). Real-Time Big Data Stream Analytics. In SIMBig (pp.13-14).

Black, D., & Thomas, J. (2013). How Gartner Evaluates Vendors and Markets in Magic Quadrants and MarketScopes. Published in 26 July 2013. Retrieved from: https://www.gartner.com/doc/2560415/gartner-evaluates-vendors-markets-magic#a-356148118.

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4),77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3,993-1022.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. Journal of Hospitality Marketing & Management, 26(7),675-693.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4),1165-1188.

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275,314-347.

Fang, B., & Zhang, P. (2016). Big data in finance. In Big Data Concepts, Theories, and Applications (pp.391-412). Springer, Cham.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2),137-144.

Gartner IT Glossary (2017). Retrieved from: http://www.gartner.com/it-glossary.

Gartner Magic Quadrant for Advanced Analytics (2017). Retrieved from https://www.kdnuggets.com/2017/02/gartner-2017-mq-data-science-platforms-gainers-losers.html.

Goes, P. B. (2014). Editor's comments: big data and IS research. MIS Quarterly, 38(3),iii-viii.

Graves, D. C., Wendel, A. A., Troescher, A. W., & Livingston, A. J. (2015). Analysis of cloud storage providers for enterprise readiness based on usage patterns and local resources. In SoutheastCon 2015 (pp.1-6). IEEE.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. Information Systems, 47,98-115.

Ittoo, A., Nguyen, L. M., & van den Bosch, A. (2016). Text analytics in industry: Challenges, desiderata and trends. Computers in Industry, 78,96-107.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. Journal of Business Research, 70,338-345.

Kretzer, M., Maedche, A., & Gass, O. (2014). Barriers to BI&A Generativity: Which Factors impede Stable BI&A Platforms from Enabling Organizational Agility? Twentieth Americas Conference on Information Systems, Savannah,20.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6,70.

Lapalme, J., Gerber, A., Van der Merwe, A., Zachman, J., De Vries, M., & Hinkelmann, K. (2016). Exploring the future of enterprise architecture: A Zachman perspective. Computers in Industry, 79,103-113.

Meng, X. F., & Ci, X. (2013). Big data management: concepts, techniques and challenges. Journal of Computer Research and Development, 50(1),146-169.

Moro, S., Rita, P., & Cortez, P. (2017). A text mining approach to analyzing Annals literature. Annals of Tourism Research, 66,208-210.

Moro, S., & Rita, P. (2018). Brand strategies in social media in hospitality and tourism. International Journal of Contemporary Hospitality Management, 30(1),343-364.

Oliveira, N., Cortez, P., & Areal, N. (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. Decision Support Systems, 85,62-73.

Romero, D., & Vernadat, F. (2016). Enterprise information systems state of the art: past, present and future trends. Computers in Industry, 79,3-13.

Salazar, R. (2014). Windows Azure. Microsoft's OS for Cloud, Web Applications, Big Data, Virtualization, Databases and Business Intelligence. CreateSpace Independent Publishing Platform, USA.

Seddon, J. J., & Currie, W. L. (2017). A model for unpacking big data analytics in high-frequency trading. Journal of Business Research, 70,300-307.

Sharda, R., Delen, D., & Turban, E. (2018). Business Intelligence, Analytics and Data Science: A Managerial Perspective (4th edition). Pearson.

Salton, G., Singhal, A., Mitra, M., & Buckley, C. (1997). Automatic text structuring and summarization. Information Processing & Management, 33(2),193-207.

Yuhanna, N., & Hopkins, B. (2016). Big Data TechRadar: A Rapidly Expanding Ecosystem. In Big Data Is Critical Technology For Insights-Driven Businesses. Forrester Research Inc., 14-31.