IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN: 2278-487X, p-ISSN: 2319-7668. Volume 22, Issue 1. Ser. VII (January. 2020), PP 20-25
www.iosrjournals.org

Big Data Analytics: The businesses future

Marwa Rabe Mohamed Ali[1]
October University for Modern Sciences and Arts (MSA)
Corresponding Author:Marwa Rabe Mohamed Ali

Abstract: *A business applying Big Data would have invested significant resources to collect, process, prepare, and eventually analyze it and accordingly expects more intense insights and knowledge as results.The Big Data concept evoluation has inspired me to review selected types ofBig Data Analytics to clarify it and highlight its expected contributions.*

Keywords: *Big Data Analytics, Customer performance analytics, Employee performance analytics*

Date of Submission: 13-01-2020 Date of Acceptance: 29-01-2020

I. Introduction

For the past 50 years, businesses and organizationshave stored and analyzed their data using relational databases and data warehouses. Although other data stores and technologies exist, the main proportion of business data can be found in these traditional systems. Nevertheless, these systems, which designed to deal with structured data, were not intended to address some of today's data challenges. Moreover, its cost would be extremely high. Importantly, sometimes data scientists spend 80% of their time in search of the right data and 20% of the time doing analytics. A data-driven environment must have data scientists spending more time doing analytics.[2]

Michael Cox and David Ellsworth used the term "Big Data" for the first time in 1997 in a paper in the IEEE conference to explain the visualization of data and the challenges created by it in computer systems *(Cox and Ellsworth 1997).*

Tech America Foundation's Federal Big Data Commission (2012) defined Big Data as a term to describe the large volumes of data, which are different, complex, variable, require techniques and technologies to acquire, store, distribute, manage and analyze information.

BD can be described as a new type of data that needs different tools and technologies to deal with, and BDA is the methods used to create insight out of it. Mostly showing up in computer literature, several Big Data definitions are focused on size and scale, others have concentrated on the technological implications - For example; McKinsey defines BD as the datasets whose size represents a challenge for traditional computing technologies*(Manyika, Chui et al. 2011). Dumbill (2012)*Have also suggested that the term applies to the data that cannot be processed using traditional tools. Therefore this paper presents an overview of Big Data Analytics and highlight some types of it.

[1] Marwa rabe is Lecturer at Faculty of Management Sciences, Accounting department, October University for Modern Sciences and Arts (MSA), Egypt. Marwa received a PhD from Cairo University in Egypt (major Accounting). Marwa is interested in Big Data Analatics, Business Analatics, and performance evaluations.

[2]http://www.pearsonitcertification.com/articles/article.aspx?p=2427073&seqNum=2

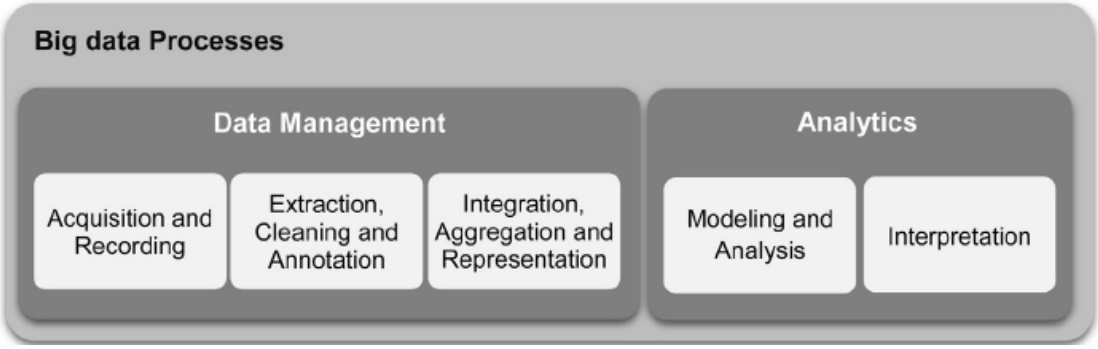
II. Big Data Analytics

Analytics is about data and how we can use it to improve business success and performance. This concept is not new, business leaders and senior executives have been using past performance and business data for decades to help decide strategy and alter it when necessary. However, what is new is the definition of what data is and the technological advances that allow us to store analyze and extract value from data that was previously impossible.

The raw material of this insight extraction process is data – whether that is traditional data or 'Big Data'. Currently the term 'Big Data' is used to describe the fact that everything we do, say, write, visit or buy leaves a digital trace, or it soon will. The easiest way to think about business analytics is that it is the process by which you take the raw material (data) and convert it into commercially relevant insights (analytics) that can inform business, improve performance and guide strategy (business intelligence).

BD can originate from traditional transaction systems as well as from new unstructured sources such as emails, audio files, social media, internet click streams, news media, videos, sensor recordings, and RFID tags *(Zhang, Yang et al. 2015).* Its potential value is released only when leveraged to drive decision-making. To enable such evidence-based decision-making, organizations need efficient processes to turn high volumes of speedy and different data into meaningful insights. The overall process of extracting insights from BD can be divided into five stages, shown in Fig 1. These five stages compose the two main sub-processes: data management and analytics. Data management contains processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis. Analytics, on the other hand, refers to techniques used to analyze and obtain intelligence from BD. Therefore, BDA can be viewed as a sub-process in the comprehensive process of "insight extraction" from BD*(Gandomi and Haider 2015).*

Fig.1 Processes for extracting insights from BD.



Appelbaum, Kogan et al. (*2017)*argued that traditionally, business and accounting data reported transactions and other structured data, such as sales, purchase orders, receivables, shipments, personal information, timesheets, and inventory. This data is organized, predictable, and familiar to businesses. This type of data stands in contrast to BD. Where the first data was structured in rows and columns, the latter data that is not structured and may seem massive to work with due to the volume, variety and data type. The appearance of BD has changed the management accountant's task. A business applying BD would have invested significant resources to collect, process, prepare, and eventually analyze it and accordingly expects more intense insights and knowledge as results.

1.2.1 Customer performance analytics

Customer satisfaction analysis

Identified as a process of evaluating whether the customers are getting what they want and expect from the business, product, or service *(Marr 2016).* Measuring customer satisfaction is a joint form of business analysis that companies engage in beyond financial analysis. It allows to find out exactly the parts of the product or service are most valued by the customers. Businesses have got into financial trouble because of imprecise assumptions about their customer's needs. Six types of analytics (Voice analytics, Video analytics, Text analytics, Sentiment analysis, Social media analysis. Web analytics) will be illustrated below to understand its mechanism.

1. Voice analytics

Voice analytics (speech analytics) is the process of extracting information, meaning and insights from audio recordings of conversations*(Marr 2016).* This form of analytics can analyze the topics or actual words and phrases being discussed in a conversation. Voice analytics can be used to extract value from what is being said and how it is being said in a way that simply was not possible a decade ago. There are many social and

commercial applications for voice analytics because it can help us to identify when someone is stressed, scared, happy, sad, or even when they are lying.

This type of analytics is valuable in helping to identify underperforming customer service representatives so they can receive additional training or coaching, and can automatically monitor the level of customer service provided on calls.*(Marr 2016).*

Gandomi and Haider (2015) reported that speech analytics follows two common technological approaches: the transcript-based approach (widely known as large-vocabulary continuous speech recognition, LVCSR) and the phonetic-based approach.

LVCSR systems follow a two-phase process: indexing and searching. In the first phase, they try to transcribe the speech content of the audio. This is performed using automatic speech recognition (ASR) algorithms that match sounds to words. The words are identified based on a predefined dictionary. If the system fails to find the exact word in the dictionary, it returns the most similar one. The output of the system is a searchable index file that contains information about the sequence of the words spoken in the speech. In the second phase, standard text-based methods are used to find the search term in the index file. Phonetic-based systems work with sounds or phonemes. Phonemes are the perceptually distinct units of sound in a specified language that distinguish one word from another (e.g., the phonemes/k/and/b/differentiate the meanings of "cat" and "bat"). Phonetic-based systems also comprise of two phases: phonetic indexing and searching. In the first phase, the system translates the input speech into a sequence of phonemes. This is in contrast to LVCSR systems where the speech is converted into a sequence of words. In the second phase, the system searches the output of the first phase for the phonetic representation of the search terms.

2. Text analytics

Text analytics (text mining) is a process of extracting value from huge quantities of unstructured text data *(Marr 2016).* Most businesses have a massive amount of text-based data from company documents, emails, reports, media releases, customer records, blogs, websites, and social media posts. While the text is structured to make sense of a human being, it is unstructured from an analytics perspective because it does not fit precisely into a relational database or rows and columns of a spreadsheet.*Chung (2014)*Argued that text analytics allow businesses to convert large volumes of human-generated text into meaningful summaries, which support evidence-based decision-making. For example, text analytics can be used to predict the stock market based on information extracted from the financial news.

Mainly, there are five main text analytics tasks: Text categorization; Text clustering; Concept extraction; Sentiment analysis; Document summarization.

Text categorization: Text analytics assigns a document to one or more classes or categories according to the subject or according to other attributes such as document type, author, creation date, etc. Text categorization applies some structure to the text, which can then be used for analysis or query. Thiscould be helpful if there were a huge amount of text data that needs to be classified for easier access and usability. Spam filters use text classification to assess the text within incoming emails and decide if the email is legitimate or not. Email routing also uses this technique to reroute an email arriving at a general address to a more appropriate recipient based on the topic discussed in the text of the email.

Text clustering allows clustering automatically huge amounts of text into meaningful topics or categories for fast information retrieval or filtering. Search engines use text clustering to deliver meaningful search results.

Concept extraction is particularly useful if there were a great deal of data that needed to access but need to do quickly to deliver results. **Sentiment analysis** is particularly useful to discover trends, patterns and hidden consensus within text over and above what the text actually says. Finally, **data summarization** allows to automatically summarizing documents using a computer program to retain the most important points from the original document. This can be useful in the case of much reading to get through but not enough time. Search engines also use this technology to summarize websites on result listings *(Chung 2014).*

3. Video analytics

*Gandomi and Haider (2015)*explained video analytics (video content analysis VCA) as "a group of techniques to monitor, analyze, and extract meaningful information from video streams". Although video analytics is still in its initial stages compared to other types of data mining, various techniques have been developed for processing real-time as well as pre-recorded videos. The increasing diffusion of closed-circuit television (CCTV) cameras and the growing popularity of video-sharing websites are the two leading sponsors to the evolution of computerized video analysis. A key challenge is the hugeness of video data where one second of a high-definition video, in terms of size, is equivalent to over 2000 pages of text *(Manyika, Chui et al. 2011). Gandomi and Haider (2015*) argued that BD technologies can turn this challenge into an opportunity. Avoiding the need for cost-intensive and risk-prone manual processing, it can be leveraged to automatically select through extract intelligence from thousands of hours of video.

Specifically, video analytics can effectively and efficiently perform surveillance functions such as discovering breaches of restricted zones, identifying objects removed or left unattended, detecting loitering in a specific area, recognizing suspicious activities, and detecting camera tampering, to name a few. Upon detection of a threat, the surveillance system may notify security personnel in real time or elicit an automatic action (e.g., lock doors, sound alarm, or turn on lights).

The data generated by CCTV cameras can be extracted for business intelligence. For example, smart algorithms can gather demographic information about customers, such as gender, and age. Likewise, retailers can count the number of customers, measure the time they stay in the store, detect their movement patterns, measure their dwell time in different areas, and monitor queues in real time. Valuable insights can be acquired by linking this information with customer demographics to drive decisions for product placement, price, and assortment optimization, promotion design, cross-selling, layout optimization, and staffing *(Gandomi and Haider 2015*).

4. Sentiment analysis

The large volume of data that has grown with the diffusion of social media platforms, the qualitative manual analysis of consumers sentiment carried in online brand-related content is no longer practical. Twitter generates more than 500 million tweets each day, and there are 4.75 billion pieces of content per day on Facebook. This raises the requisite to develop automated techniques for identifying and analyzing consumer sentiment expressed in text *(Wang, Chen et al. 2012).*

Sentiment analysis (opinion mining) search for extracting subjective opinion or sentiment from video, text, or audio data. The main aim of it is to determine the attitude of an individual or group concerning a specific topic or overall context. The sentiment or attitude may be a judgment, evaluation, or emotional reaction. Obviously, what someone thinks and feels is very subjective so the data you have toanalyze this subjective element would need to indicate sentiment in some measurable way. As a result, sentiment analysis can be applied to text, speech (audio) and visual interactions (video).

Qazi, Raj et al. (2017) argued that sentiment analysis aims to elicit opinions about entities (such as products or services) to get useful information. Furthermore, its goal is to present the information in such a way that serves the objective of both customers and manufacturers. It demonstrated that many potential consumers and manufacturers overlook detailed reviews, preferring summarized opinions. The summary is based on the concept of classification of opinions, identifies whether opinions hold negative, positive, or neutral sentiments

*Prabowo and Thelwall (2009)*explainedthatby increasing the volume of information on the internet, it becomes increasingly tough and time-consuming to make decisions about the purchase of products and services. Opinion and sentiment analysis (SA)are techniques that aim to automate the analysis of information and thereby save the user time and effort. It can be viewed as an emerging area of significance in the computing discipline, relying on a mixture of new and established methods. *Prabowo & Thelwall*gave an example, in the sentence, "My doctor was great, but the staff was horrific." "The word," "horrific," is used to categorize the second part of the sentence as a complaint and the term, "staff," is used to classify the comment as related to the staff.

The website, www.Ratemds.com, which contains patient satisfaction data on many different physicians. The site allows patients to rank their doctors' staff, punctuality, helpfulness, and knowledge on a scale of 1-5; a ranking of 1 corresponds to the phrase, "did not meet my expectations," while a score of 5 indicates that the doctor, "exceeded my expectations." Also, the questionnaire contains a comments section that enables the patients to discuss the particular aspects of their care, which they liked or disliked *(Hopper and Uriyo 2015).* Reading through each patient comment and marking them either as a complaint or as praise, based on the nature of the text. In each case, the basic sentiment analysis system can be mimic by using keywords (e.g. "love," "like," "awful") and word groupings to guide our decisions. The total number of positive and negative keywords were counted and groupings. A comment is considered to be a compliment if it was mostly positive. Otherwise, it treated as a complaint.

*Marr (2015)*statedsome software tools (social mention, Twitter Sentiment, Yacktracker and Twitrratr) that can help in measuring the sentiment around the product or service. Twitrratr, for example, can separate the positive tweets about the company, product, brand, or service from the negative and neutral.

5-Web analytics

Web analytics is the process of analyzing online behaviour to optimizewebsite use and increase engagement and sales. There are two types of web analytics – off-site and on-site. Off-site web analytics looks at what is happening on the internet as a whole and includes the measurement of a product or service's potential audience, competition, and online trends. On-site web analytics is the analysis of your website. This includes collecting data on the number of people visited the site, where they came from, how long they stayed, how they navigated the site and whether the visit resulted in a sale. Off-site web analytics is beneficial for evaluating the market and opportunity whereas on-site is useful to measure financial results *(Marr 2016).*

Google analytics is probably the front-runner among web analytics tools and service providers. What is brilliant about these tools is that the possibility to set up what you want to measure and asses and the tool will do all the work. There are also tools such as CrazyEgg, which shows what parts of the website are 'hot', warm, or 'cold'. Hot and warm areas of the site indicate where customers are visiting and staying. Cold areas are where there is no traffic or the customer quickly leaves. These insights can help to refine the online presence and give the customers more of what they demonstrate, they want and less of what they show they do not want.

6-Social media analytics

Social media analytics refers to "the analysis of structured and unstructured data from social media channels". Social media is a term covering a variety of online platforms that permit users to create and exchange content. It can be classified into the following types: Social networks (e.g., Facebook and LinkedIn), blogs (e.g., Blogger and WordPress), microblogs (e.g.,Twitter and Tumblr), social news (e.g., Digg and Reddit), social bookmarking (e.g., Delicious and Stumble Upon), media sharing(e.g., Instagram and YouTube), wikis (e.g., Wikipedia and Wikihow),question-and-answer sites (e.g., Yahoo! Answers and Ask.com) and review sites (e.g., Yelp, TripAdvisor) *(Barbier and Liu 2011).*Moreover, many mobile apps, such as "Find My Friend", provide a platform for social interactions and, therefore, serve as social media channels.

*Gandomi and Haider (2015)*reported the user-generated content (e.g., sentiments, videos, images, and bookmarks) and the relationships and interactions between the network entities (e.g., people, organizations, and products) are the two sources of information in social media. Thus, the social media analytics can be categorized into two groups:

Content-based analytics: focuses on the data posted by users on social media platforms, such as customer feedback, product reviews, images, and videos. Such content on social media is often voluminous, unstructured, noisy, and dynamic. Text, audio, and video analytics, as discussed earlier, can be applied to derive insight from such data.

Structure-based analytics (social network analytics) this type of analytics are concerned with synthesizing the structural attributes of a social network and extracting intelligence from the relationships among the participating entities. The structure of a social network is modelled through a set of nodes and edges, representing participants and relationships, respectively.

1.2.2 Employee performance analytics

Achieving company growth and improving profitability need to develop and motivate employees to maintain the skill level required and to ensure keeping talented employees. Employee performance analytics search for assessing individual employee performance. The resulting insights can identify who is performing well and who may need some additional training or support. Moreover, a grasp of employee performance can also contribute to the recruitment process, thus the accurate types of employees are recruited, and the costly mistakes are avoided*(Marr 2016).*

Nowadays, there are many innovative ways of collecting and analyzing performance. The data collected about employee performance can, however, be further analyzed to provide additional insights. They can, for example, be recorded which can then be analyzed using voice analysis, or converted to text for text analysis, or data mining.

1. Sociometric Solution

The key concept at the core of Sociometric Solution is the "human cue": each person spreads around many cues that can be captured through voice and position in space, offering behavioural signals ready to be interpreted. The Sociometric Badges are small sensor-rich tools that employees wear for a certain time, allowing the gathering and tracking of a massive amount of data related to a set of human cues. Tracking includes tone of voice; time spent speaking versus listening, connections and exchanges among individuals, movement and routes around the offices. All data collected in this way is analyzedto understand the human dynamics in the organization, through quantitative and objective data representing, as much as possible, the reality. The kind of insights offered by Sociometric enables the testing of hypotheses on layout, organization and team structure that analytics intensive companies usually perform on their websites and digital platforms, allowing to modify organizationsto increase effective communication and interaction between employees*(Morabito 2015).*

2. Capacity analytics

Capacity analytics allow tracking how much time the employees spend doing their work and how much time they spend doing other work through the video analytics and social media analytics. It seeks to establish how operationally efficient individual employees are to establish capacity levels. In addition, it allows the company to understand how much capacity they have to take on new projects. If everyone is at 100 percent capacity then taking on any more work is not advisable unless capacity can be increased by, for example,

recruiting new staff. Moreover, identify patterns and trends in employee performance that can then be used to improve recruitment or training and development *(Marr 2016).*

3. Employee churn analytics

Employee churn analytics is the process of assessing the staff turnover rates in the past in an attempt to predict the future.Its importance is because of hiring and training employees costs time and money. When that investment is lost because too many employees are leaving the business, then this can have a detrimental impact on the business *(Marr 2016).* Historical employee churn can be identified through traditional key performance indicators such as the employee satisfaction index (ESI), employee engagement level and staff advocacy score. Likewise, the use of surveys can help to gather additional information that can then be mined for greater insights. Text analytics can be used on that data as well as performance reviews and social media data.

The aftermentioned analytics dealing mainly with human data, which seems to be relevant to measure customers and employees performance in the modern business environment.

References

[1]. Appelbaum, D., A. Kogan, et al. (2017). "Impact of business analytics and enterprise systems on managerial accounting." International Journal of Accounting Information Systems**25**: 29-44.

[2]. Chen, J., Y. Tao, et al. (2015). "Big data based fraud risk management at Alibaba." The Journal of Finance and Data Science**1**(1): 1-10.

[3]. Cox, M. and D. Ellsworth (1997). Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th conference on Visualization'97, IEEE Computer Society Press.

[4]. Dumbill, E. (2012). "What is big data? An introduction to the big data landscape." oreilly. com, http://radar. oreilly. com/2012/01/what-is-big-data. html.

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

[6]. Hopper, A. M. and M. Uriyo (2015). "Using sentiment analysis to review patient satisfaction data located on the internet." Journal of health organization and management**29**(2): 221-233.

[7]. Marr, B. (2015). Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance, John Wiley & Sons.

[8]. Marr, B. (2016). Key Business Analytics: The 60+ Business Analysis Tools Every Manager Needs to Know, Pearson UK.

[9]. Manyika, J., M. Chui, et al. (2011). "Big data: The next frontier for innovation, competition, and productivity."

[10]. Morabito, V. (2015). Big Data and Analytics: Strategic and Organizational Impacts Springer International Publishing Switzerland, ISBN 978-3-319-10665-6, DOI 10.1007/978-3-319-10665-6_1.

[11]. Qazi, A., R. G. Raj, et al. (2017). "A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges." Internet Research**27**(3): 608-630.

[12]. Prabowo, R. and M. Thelwall (2009). "Sentiment analysis: A combined approach." Journal of Informetrics**3**(2): 143-157.

[13]. Wang, W., L. Chen, et al. (2012). Harnessing twitter" big data" for automatic emotion identification. Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), IEEE.

[14]. Zhang, J., X. Yang, et al. (2015). "Toward effective big data analysis in continuous auditing." Accounting Horizons**29**(2): 469-476.

Marwa Rabe Mohamed Ali. "Big Data Analytics: The businesses future." *IOSR Journal of Business and Management (IOSR-JBM)*, 22(1), 2020, pp. 20-25.