BIG DATA ANALYTICS: TO UNDERSTAND THE CHALLENGES

AUTHORS

DHATRI MAHESHWARI

Post Graduate Management Student, Universal Business School, Karjat, Raigad, Maharashtra, 410201. Email address:dhatri.maheshwari@ubs.org.in

JENISHA UPADHYAY

Post Graduate Management Student, Universal Business School, Karjat, Raigad, Maharashtra, 410201. Email address:jenisha.upadhyay@ubs.org.in

DESHNA NAYAK

Post Graduate Management Student, Universal Business School, Karjat, Raigad, Maharashtra, 410201. Email address:deshna.nayak@ubs.org.in

SWARAJ AWASTHI

Post Graduate Management Student, Universal Business School, Karjat, Raigad, Maharashtra, 410201. Email address:swaraj.awasthi@ubs.org.in

VIRAJ RANDIVE

Post Graduate Management Student, Universal Business School, Karjat, Raigad, Maharashtra, 410201. Email address:viraj.randive@ubs.org.in

PROF.M.GURUPRASAD

HOD -General Management, Dy.Director – Research, Email address: mguruprasad@universalai.in

ABSTRACT

The present contemporary age has been appropriately identified as ‘an era of uniform data or alsothe systematic monitoring of every citizen’s way of the communications channels or the actionsof the users through the use of the information technology. It is also the era that is called the eraof big data or also the processing of such huge and enormous amounts of information. These all-data sets can be easily looked upon or searched upon, can also have attached links to other data sets, and can be differentiated through their overall comprehensive reach and long-term collection. The only challenge that comes with big data is the privacy problems that come from boththe state and the non-state actors present.

KEYWORDS

Technology, Big Data, Analytics, Data, Artificial Intelligence, Privacy

INTRODUCTION

The process of uncovering hidden trends or the patterns in the data, and associations of the data in the large volumes which is the raw data in order to make more data- informed and beneficial decisions which can help to improve the business effectively, this process is known as big data analytics. This has been benefiting to the companies for very long but as said everything has its dark side and the research here is done to bring out the darker side of the big data analytics. Thepaper covers all the possible issues which are attached with the big data, and some of the real lifeexamples like privacy issues, or storage issue which need to be solved before adapting to more advanced technological world.

Big data analytics is an advanced way of analyzing enormous data sets which vary from structured, non-structured, or semi-structured data from various sources. It is often referred to asa complex process because it requires a deep examination of big data through which the hidden patterns, recent or future trends, business strategy, and customer preferences can help organizations and businesses to make more informed and effective decisions. The process of uncovering hidden trends or patterns in the data, and associations of the data in large volumes which is the raw data to make more data- informed and beneficial decisions that help to improvethe business effectively, this process is known as big data analytics. This has been beneficial to the companies for a very long but as said everything has its dark side and the research here is doneto bring out the darker side of big data analytics. The paper covers all the possible issues which are attached to big data, and some of the real-life examples like privacy issues, or storage issues which need to

be solved before adapting to a more advanced technological world.

WHAT IS BIG DATA ANALYTICS?

Big data analytics is an advanced way for analyzing enormous data sets which varies from structured, non-structured or semi structured data from various sources. It is often referred to as a complex process because it requires deep examination of big data through which the hidden patterns, recent or future trends, business strategy and customer preferences can help organizations and businesses to make more informed and effective decisions.

It also involves some analytic models through which the data is analyzed through predictive models and some algorithms of statistics.

This data analytics software can use this to benefit the business in terms of efficiency with reference to operations, and can have competitive advantage over the competitors.

HOW DOES BIG DATA ANALYTICS WORK?

All the Data analysts, scientists, the people who are the predictive modelers and other professionals firstly collect, then process, clean the data which is processed and analyses it to very depth to bring out the important information from it.

Here is a brief of the four steps which are involved in the data preparation process:

The data professionals collect the data from different sources which have variations, but most of the time it is a mix of unstructured and semi-structured data. Most of the organizations use different types of stream to get the data are:

- Web servers log:
- Clickstream data
- Mobile and cloud applications
- Content posted on social media.
- Survey responses
- Emails and text messages from customers
- Phone records
- Machine data

Further the data is processed after which the data is stored in the warehouse or data lake after which the data is organized by the professionals then happens the configuration and then the datais partitioned into

different analytical queries, as the data is processed there are higher chances ofgood performance of the analytics.

RESEARCH OBJECTIVES

1) To understand evolution of Big Data Analytics

2) To understand the challenges due to Big Data Analytics

METHODLOGY

This research paper employs a comprehensive secondary research methodology, drawing insights from existing literature, academic publications, and relevant sources to explore the evolution and darker side of big data analytics. The systematic literature review focuses on synthesizing and analyzing information from diverse secondary sources to achieve the research objectives

LITERATURE REVIEW

The research explores "the dark side of big data analytics" using a systematic literature review methodology. The primary data source is EBSCO, chosen for its global scope. The keywords "Big data" and "Technology" were initially employed, expanded with synonyms for relevance. This reflects a comprehensive approach to understanding the implications of big data analytics. (EBSCOhost Research Platform: EBSCO). (Jessica, year)

The literature review identifies negative aspects of big data analytics, emphasizing its impact on society and privacy. Jessica's work explores the societal consequences of modern technology, emphasizing privacy concerns due to excessive use. Riva underscores big data's significant effects on science and society, cautioning against the portrayal of its fusion with AI as a universal solution. (Big Data is Watching YOU: opportunities and Challenges from the Perspective of young adult consumers in Russia.) (Zero Sales Resistance: The Dark Side of Big Data and Artificial Intelligence.) (Jessica, year)

The dark side of new technologies, such as AI, block chain, and big data analytics, in B2B partnerships is discussed. The study acknowledges strides in understanding negative effects but suggests an undertheorized perspective. It explores how emerging technologies may disrupt partnerships, creating dark side effects yet to be fully understood. (Theorizing the dark side of business-to-business relationships in the era of AI, big data, and block chain.)

The intersection of AI and big data analytics in ridesharing user engagement is explored, considering both positive and negative effects. The study delves into how these technologies affect users' cognitive assessments and alter their participation. It contributes to information security research by revealing the interplay between positive and negative aspects of BDA and AI. (The good, the bad, and the ugly: impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing. (Riva, year)

The review identifies a growing digital divide caused by the rise of big data. Small and medium-sized businesses face challenges in accessing and utilizing big data, leading to potential monopolies by larger corporations. The consequences of this pattern are discussed, highlighting issues such as lack of standard measurements and variations in big data impact across industries. (Big Data and Analytics: Issues, Solutions, and ROI.) The subsequent sections of the research paper are outlined, covering aspects from introducing big data analytics to presenting real-time examples and recommendations.

The present contemporary age has been appropriately identified as ‘an era of uniform data or alsothe systematic monitoring of every citizen’s way of the communications channels or the actionsof the users through the using of the information technology. Itis also the era which is called as the era of big data or also the processing of such huge and enormous amounts of the information.These all data sets can be easily looked upon or searched upon, can also have attached links to other data sets, and can be differentiated through their overall comprehensive reach and long- term collection. The only challenge that comes with big data is the privacy problems which comes from both the state and of the non-state actors present. (Riva, year)

To be very sure, we have an example of the Aadhaar Card, which is also considered as the Big Data, because it is the billed as the world's biggest known as the biometric project. Every Indianwho resides here is provided with a unique and personalized 12-digit digital number under the Aadhaar, which will not only include all the basic personal data like the name, address, and dateof birth, but will also have the details about both the ten fingers along with the iris are scanned. And stored in the system.

All of this information is held in a safe vault in Manesar, Haryana. The project had issued approximately 1.171 billion cards as of August 15 this year, at a cost of over Rs 9,000 crore. It'slikely that the Supreme Court would rule that Aadhaar does not violate the right to privacy. It may try to strike a balance between public interest and privacy by instructing the state to enforcea strong data protection law, as some legal minds have suggested. Despite this, the government'salmost fanatical apology for Aadhaar, as well as its

argument against privacy being a fundamentalright.

EVOLUTION OF BIG DATA ANALYTICS

Early Stage: Emergence of Big Data Concept

The concept of Big Data emerged with the recognition that traditional data processing tools were inadequate to handle the vast volumes of data being generated. Doug Laney's early articulation of the three V's—Volume, Velocity, and Variety—in 2001 laid the foundation for understanding the unique challenges posed by massive datasets (Laney, 2001).

Mid-2000s: Technology and Frameworks

The mid-2000s witnessed the emergence of technologies and frameworks that laid the groundwork for Big Data Analytics. Apache Hadoop, an open-source framework for distributed storage and processing, became a cornerstone. The release of the MapReduce programming model by Google and its subsequent implementation in Hadoop allowed organizations to process large datasets in parallel across clusters (Dean & Ghemawat, 2004).

Late 2000s to Early 2010s: Mainstream Adoption

Big Data Analytics started gaining mainstream adoption during this period. Organizations recognized the potential of extracting valuable insights from large datasets for informed decision-making. Technologies like Apache Spark, which provided faster and more flexible data processing, gained prominence, complementing Hadoop's batch processing capabilities (Zaharia et al., 2010).

Mid-2010s to Present: Advanced Analytics and Machine Learning

The evolution of Big Data Analytics transitioned towards more advanced techniques, including machine learning and predictive analytics. The integration of machine learning algorithms into analytics platforms allowed organizations to uncover patterns, trends, and correlations in data for predictive modeling (Chen et al., 2018).

Real-Time Analytics and Cloud Computing

Recent years have seen a focus on real-time analytics, facilitated by technologies like Apache Flink and advancements in cloud computing. Cloud-based platforms, such as Amazon Web Services (AWS) and Microsoft Azure, provide scalable infrastructure for storing, processing, and analyzing large datasets in real-time, making Big Data Analytics more accessible and cost-effective (Vaquero et al., 2018).

Future Trends: AI Integration and Edge Computing

The evolution of Big Data Analytics continues with a trajectory toward greater integration with artificial intelligence (AI) and edge computing. AI techniques, including natural language processing and computer vision, are increasingly being applied to derive deeper insights from diverse datasets. Edge computing allows data processing to occur closer to the source, reducing latency and enhancing real-time analytics capabilities (Zhang et al., 2019).

THE ISSUES AROUND BIG DATA ANALYTICS:

GARBAGE IN PROBLEM

With the volume of data increasing day to day, the real challenge is to determine how to derive the values and insights from such a large amount of the data. The real issue with the unstructuredand semi-structured data is the quality which will have to be addressed, for which more scalable,more reliable and high end technological resources are needed which is handled by some extremely smart, professional and experienced Data scientists who could give the surety about the data quality and can handle and solve all the complex problems.

OPEN MISUSE OF DATA

The data scrapers of the data analysis collect the keywords, identity like names, and the content categories of which the data has to be collected from a client's website. Web crawling is beneficial to the companies through which the keywords are used to drive the traffic of the website, the content categories are created to get the user attention, and the titles are ranked so that they are on the higher side at the SEO and various other uses. The collected data insights help people to get the monetary benefits. However everything has a dark side and here it is the open misuse of the data by the companies. This data is misused for the various purposes, like undisclosed information to be gathered or to share it with third parties illegally to get the monetary benefits to generate the revenue. This not only limits personal misuse but also professional benefits.

THE END OF PRIVACY

It has been a difficult phase for the world to transact if we don't keep a check on a regular basis.Most of the people around are in worries about their data being stored and used. Of course, nobodydislikes getting promotional deals, discount offers or coupons but everything has its ownlimitations, the line is crossed when the offering company follows continuously and tracks all your transactions. This recruiters search for the data for any employable purposes using data mining, which is a dark side in terms of HR force.

Companies don't want an environment where stakeholders get involved in stalking a prospect anywhere in an online way. The companies must be aware about the fine line which exists between providing the stakeholders and stalking about various advice or solutions. Also, the business must need to assume responsibility of providing a positive experience to all the customers of the company instead of stalking them.

Apart from corporations being rogue with the big data, as it creates chaos for the businesses, which sometimes are uncontrollable. The Data Crawling and its service providers, such as PromptCloud, would be helpful here. The following are some of the most common issues that businessescan face:

DATA INTEGRATION PROBLEM

Big data adds up to the big mountain of complete unstructured data that can render the data integration a very difficult job. Big data in the enterprise comes through various sources, including all the transactional databases available, various mainframes and the ERP systems, as well as from all the customers and suppliers. Moreover, these available sources do have their ownmodel of the data, which exacerbates the problem. The available enterprise data, which is equallydistributed and completely decentralized, also adds up to the difficulty of the situation by makingthe data integration much more difficult. To be able to extract the value from business data is a very time heavy and difficult task to execute.

UNSTRUCTURED DATA

With all the available data which is being produced or getting extracted from various sources in every second, we already have a large volume of the unstructured data which is huge to interpret.All the businesses want to address the various business issues by continuously extracting useful data which could give insights from the available unstructured data sets. Companies will have toaddress the data integration problems, data consistency, and the total unstructured text analyticsproblems by referring to the business decisions they want to make in the future.

RELEVANCE

In the world which is full of information technology, Big Data analytics is the true revolution. Each and every year, the use of data analytics by any businesses grows continuously. Big data has a lot of big range, length, and velocity to it. All the Machine learning, data analysis, natural language processing, and statistics are some of the analytics methods used in Big Data.
There are multiple operations that can be conducted on a single platform using of big the data. By using of a few big data resources, you can be able to store terabytes of data, pre-process the data, and can analyses it along with the visualization. (How to Overcome the Dark Side of Big Data Analytics).

To provide all the analysis for companies, the data is extracted and then prepared, after which itis blended. Many large corporations and the multinational corporations use these tactics in a various different forms these days. (Hussain, What is Big Data & why is Big Data important in today's era 2019).

Big data is a concept that has been continuously thrown around a lot in the present tech world inthe recent years. Essentially, to which the amount of data which the companies are able to mine for information or the insights is rising rapidly each and every year, and for that we should thankto modern software tools, it is also possible to analyses and use this data to improve business efficiency. What role will the big data play for being the driving change for HR professionals? Big data has a lot of role to play in the Big Data Analytics which is written below in detailed.

Big data analytics also enables businesses to work more efficiently and effectively with their datawhich is available to them and to use the available data to find various potential opportunities. To be able to predict from the data, different variety of various techniques and algorithms can beused. Multiple strategic and pre planned strategies can also be implemented for the company's success which will be resulting in wiser and informed business decisions, more and more successful business operations, and very much higher profits. (About the Author: Guest Contributor Aberdeen's Guest Contributors are subject matter experts from a wide range of roleswithin various organizations. These folks are notable thought lea & Aberdeen's Guest Contributors are subject matter experts from a wide range of roles within various organizations.These folks are notable thought leaders in their fields who, *6 Ways Big Data Will Change HR*)

PROBLEMS IT WILL POSE

When everyone is navigating and understanding this new world of the big data, there are severalobstacles which are to be or required to overcome. It also necessitates the ability to be able to find, or extract, and analyses appropriate amount of data from a variety of sources. The total complexities of navigating the data after it has been processed also includes the preserving up ofthe information for the easy retrieval and sharing it while maintaining a complete privacy

These all the challenges amounts to one conclusion for HR professionals: which is investing in an effective HRIS is extremely critical in our current business environment. Once a complete transparent or satisfactory system is put into the place, the software can improve various elementsof your business, such as the lower expenses, all the productivity gains, and a very higher competitive advantage.

SOLUTIONS

If you are not experience and are new to the world of big data then it is a good idea to pursue professional guidance or assistance. To be able to take the big data consulting, you can hire an expert who is best in his field or can also switch to an available vendor. In any of such case the collaborative and combined efforts would help to enable one to develop a particular plan and onthe based on which, one can select the appropriate technology available.

Since the big data is of such importance the change for an organization, it must be adopted firstlyby all the top management of the company and then can be adapted by the lower management. The company's IT departments must arrange numerous trainings and seminars to make sure thatbig data is comprehended and accepted at all the different levels uniformly. And the introductionand use of the latest available big data solutions must be managed and maintained properly to ensure even greater big data acceptance.

There is already a very strong call being taken for the source codes used in the decision making algorithms which are to be made available to the public. At various levels, such transparency andopenness can be a lot counterproductive. It would resultantly expose or fail all the guidelines andactivities and the formula they use in the tax audits, allowing the evaders to manipulate and trickthe system. Implementing a transparent system will also link intellectual property rights, which will allow rivals to benefit from the company's proprietary algorithms, which would demotivateto bring out better algorithms.

Nobody is trying to hide the fact that big data is not perfect. And, in the end, it isn't that relevant.But that doesn't mean you shouldn't have any say on how trustworthy the data is. It can not onlycontain incorrect

facts, but it can also repeat itself and contain contradictions. And it's unlikely that data of such poor quality can provide any valuable insights or exciting opportunities.

All the stakeholders can come together and can develop specific operational frameworks for the best responsible use of the big data analytics, the given approach to the algorithm justice is for all the stakeholders, who can come together and can create a clear operational frameworks for the using the big data analytics consciously and responsibly. Testing for both the fairness and thepotential prejudice during the development stages, as well during the ongoing audits, will allow us to ensure that algorithms do not have discrimination or breaking any laws.

Reaching a common or one consensus on the side of such a stakeholder efforts, whether they should be extending to the credit, housing, jobs, and other areas of economic life, will be a majorchallenge in such an endeavor. Another difficulty is defining the positions of the various stakeholders, which include algorithm creators, consumers, regulators, independent researchers,subject matter experts, ethics experts, and the general public. For all the statistical models whichare used in credit, housing, and jobs, there already exists some rules in place that will include thecomplete validation of predictive accuracy. (Bekker, the 'Scary' Seven: big data challenges and ways to solve those 2020).

REAL LIFE CASES

PRESIDENCY OF MR.TRUMP

- Data in today's time is being continuously used to swing out the real public opinion in manypolitical areas and as the data present today are being powered through automated machines,it is possible that they could capture very big quantities of data quite easily, and which wouldbe used by the computers that can interpret and provide these data insights to the politicians for the personal benefits.

- Mr. Robert Mercer who is an American based hedge-funds billionaire who was one of the most important in the financial committee of the Trump presidency. He just recently was revealed as one of the owners of the Cambridge Analytical (CA), which is a data mining anddata analytics company that use all the available technology and data mining tools to build the voter profiles through which the political advertisement and the messages can be deliveredto the targeted emotional triggers and the votes can be swinged to a side of any politician. This technique or process through which this is done is known as “psychographic”.

- Steve Bannon who is Trump's Chief Strategist is also a member of the White House SecurityCouncil,

and also a member of the CA's BOD. This organization has strong links with the Republican Party.

 They (CA) collected around 5,000 total data points from around more than 220 million available Americans and according to the company's website, it is said that they used more than about 100 data variables to model all the audience groups and also predict their further actions. They claim that by totally aggregating the huge amount of data it gathers it can easilybuild a Big Five personality profile for all the adults in the America. The big 5 model consistsof 1)-Openness, 2)-Conscientiousness, 3)-Extraversion, 4)-Agreeableness, and 5)- Neuroticism.

 This also allows and enables Trump's campaign team to be able to personalise advertisementsfor the messages he wants to send out during his campaign.CA would also group the people who also have the similar needs and they also aspires for the similar things which could alsotogether using personality profiling, even when they are physically very different from each other.

 CA's election and voting research was so detailed that it also gave Trump's campaign an advice about where the rallies can be conducted and can be successful for the campaign, alsohe advised where to air the election campaign TV advertisements which will be able to find the potential supporters of the

 Campaign and how would it maximize all the hired volunteers' door-to-door campaign and interaction visits easy.

 In this cut to cut and fiercely contested national election which was against Hillary Clinton so it was reported that Donald Trump's election campaign roughly spent more than $6 millionto put themselves in the position to have all the necessary advantage by targeting the swing voters in the specific locations.

 Each and Every day, the Trump president campaign was using around 40-50,000 various variants of an advertisement which was constantly and continuously monitoring the responsesand after which it was adjusting itself and evolving itself on the basis of the responses whichwas stated according to - Mr. Martin Moore who was the director of the King's College's Centre which was one of the best university for the Study of Media and Communications along with Power. He also stated that everything was done by maintaining full secrecy and he could spend as much money as they wanted on the target and specific locations so that they could focus on these locations.

PRIVACY CONCERS

One of the Russian crime ring has been able to collect the largest documented number of stolen or hacked Internet passwords, including more than 1.2 billion usernames and passwords combinations and around more than 500 million email addresses,

According to the New York Times. This came after the Eastern European hackers stole around 40 million total credit card numbers from the Target and Vietnamese also the data thieves stole around 200 million total personal records which included all the Social Security numbers, all thecredit card data, and also had all the bank account information from which is now owned by thedata brokerage firm Expert.

ABUSES

Farm Bureau Federation is now calling for very stricter restrictions on about how the data provided by farmers in the area is used by companies in which they run their businesses.

According to Farm Bureau famous economist known as -Matt Erickson, he stated his concerns that the groups completely opposed to particular activities, like the increasing use of geneticallymodified organisms which will be able to gain access to big and anonymous data and links it to specific farms- just as the hackers recently linked unknown Netflix data to its specific customersand using the data against individual farmers. (Team & -, *the Dark Side of Big Data* 2019).

GOOGLE FLU TRENDS

Big Data is all about bringing or discovering any trends and similarities in the human behaviour,but that is not the case every time. One such example is Google, which also claimed that it coulduse the Big Data to be able to forecast flu outbreaks during 2008 and then they started the programmer known as Google Flu Trends (GFT). This algorithms looked for the trends in the search of the queries and enquires of containing the word "flu," but it also ignored the unrelatedsearches. After that in 2013, it collapsed drastically by missing the height of the flu season by total One Hundred and Forty percent, you make call it arrogance or a big blunder of judgement.But the truth is that the correlation does not necessarily means the causation, and often Big Datais just too large and irrelevant in several cases. (When Big Data goes Wrong 2021).

CONCLUSION

In today's extremely data-driven environment the management of the data is extremely critical and should not be overlooked. One must be very accurate and proactive in learning about the data strategies and implementing the data strategies that are aligned with one's personal or organizational objectives. One will be able to effectively mitigate any kind of the big data issuesby doing so. To handle the data and manage it properly the companies will need to recruit or collect a dedicated and focused team of professionals and experts. On the other hand the moderndata based tools can provide a very easy way to supplement and to

optimize the work of the current workers so that in order to transform the available data into the valuable market insights.Yes, the big data may have a dark side to it which can never be overlooked, but if we become more proactive, informed, expert we can handle it in very good ways which could also benefits the companies. Also if we can make better choices today than we can make the future while working with data on the brighter side by using all the expertise and leadership skills for the betterment.

REFERENCES

- *Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499 (2010)*
- *Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)*
- *Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011)*
- *Big Data Analytics: A literature review paper - researchgate. Available at: https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper).*
- *A survey on big data analytics challenges - thesai.org. Available at: https://thesai.org/Downloads/Volume7No2/Paper_67-A_Survey_on_Big_Data_Analytics_Challenges.pdf.*
- *(PDF) the spectrum of Big Data Analytics - ResearchGate. Available at: https://www.researchgate.net/publication/331054979_The_Spectrum_of_Big_Data_Analytics.*