

IMPACT OF BIG DATA ANALYTICS ON BUSINESS DECISION MAKING

Yedhukrishnan M Pillai Guide: Asst. Prof. Gauri Ansurkar

Keraleeya Samajam's Model College, Khambalpada Road, Thakurli, Dombivli (East), Kanchangaon, Maharashtra

ABSTRACT

Big Data Analytics results can play a major role in corporate decision-making allowing companies to achieve competitive advantage and make improved decisions. The study aims to investigate inhibitory and enabling factors that influence the effective implementation of Big Data (BD) and Big Data Analytics (BDA) for strategic decision-making purpose. The Big Datas are a phenomenon that is causing a progressively greater impact in different fields at a global level finding a large diffusion in business realities, influencing accounting models, control approaches, decision-making processes related to strategic planning and marketing, operational processes and new product development. Over the last 10 years most of the organizations use Big Data to improve their standards with respect to quality and cost. Big Data is a broad and mosaic set of unstructured and structured data which sizes over exabytes $\approx 10^{16}$.

A significant amount of digital data is created when the organizations convert their data from analogy to digital. The data keeps on increasing and petabytes of information are generated every year, which leads to complexity in handling data. There are many dynamic design challenges which lead to no comprehensive design strategy for Big Data. Many open sources and commercial data analysis tools are developed and are significant. Investments on Big Data have a steep hike year by year, which is a good sign in the perspective of business intelligence and decision-making capabilities of the organizations

INDEX TERM

Analytics, tools, technology, software, Data logging

1. INTRODUCTION

Big data is the name indicates deals with the humongous volume of data. Candidly talking, these chunks of data that are too thick to meet in an excel sheet. In vocational terms, handling large data using conventional database management systems is a bit tricky. Big Data needs special kind of tools to compute these massive quantities of evidence. We need even more computing to extract the full data and extra tools to study them.

Here, the large blocks of documents are accumulated from

many sources; therefore, the information is entirely bleak and in most instances it is amorphous. If we desire to use the unstructured data for an objective, it will not serve us. Data pre-processing has to be executed as an initial step and will require

just about nonconventional methods to store the information equally well. Many of us have a small confession regarding data mining and Big Data. Data mining also deals with vast quantities of data and Big Data as well. In Big Data most of the data is unstructured. In general words, we can differentiate Big Data as an “Asset” and data mining as a “handler” of the property. Over the years, big data analytics has evolved with the adoption of agile technologies and the increase of focus on advanced analytics. There is no single technology that encompasses big data analytics. Several technologies work together to help

companies procure optimum value from the information. Among them are machine learning, artificial intelligence, quantum computing, Hadoop, in-memory analytics, and predictive analytics. These technology trends are likely to spur the demand for big data analytics over the forecast period. Earlier, big data was mainly deployed by businesses that could afford the technologies and channels used to gather and analyze data. Nowadays, both large and small business enterprises are increasingly relying on big data for intelligent business insight. Thereby, they boost the demand for big data.

2. WHAT IS BIG DATA?

The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size. Any data that can be stored, accessed and processed in the form of fixed format is termed as a ‘structured’ data. Over the period of time, talent in computer science has achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also deriving valu

out of it. However, nowadays, we are foreseeing issues when a size of such data grows to a huge extent, typical sizes are being in the range of multiple zettabytes.

Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc. Now day organizations have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format.

3. BIG DATA ANALYTICS TOOLS

The rise of big data has created a huge market for data analytics tools that help enterprises seamlessly implement big data solutions. Startups like Ople and Cruz Informatics have experience working quickly with large corporations to make data ready for action.

Ople was founded with a very specific vision: Artificial Intelligence should be easy, cheap and ubiquitous. Ople accelerates the data science process, enabling organizations to attach more challenges and get answers sooner. With Ople, Data Scientists can focus on the business at hand, not the plumbing, creating 10X more production quality AI models and reducing time to deployment from months to just a few days.

Using data to uncover actionable insights is the life-blood of an industry. Crux makes data ready for action by taking on the burdensome aspects of our customers' information supply chains—helping firms acquire, explore and transform data with ease—so they can focus on what really matters. With its Informatics Platform, Crux offers a secure, scalable environment to store, explore and transform data through its integrated cloud service. Crux makes data delightful.

With the help of big data, companies aim at offering improved customer services, which can help increase profit. Enhanced customer experience is the primary goal of most companies.

4. TECHNOLOGY & SOFTWARE

Big Data technologies are the software utility designed for analyzing, processing, and extracting information from the unstructured large data which can't be handled with the traditional data processing software.

Companies required big data processing technologies to analyze the massive amount of real-time data. They use Big Data technologies to come up with Predictions to reduce the risk of failure.

1. APACHE HADOOP

Apache Hadoop is the single most characteristic software tool in the field of big data. Originally developed by Yahoo and due to its open-source license extended by Google, it is the foundation of big data in most companies. It is the software framework for the storage and processing of data. The software is extendable by a number of modules to adapt it to the needs of the user. The software framework is mostly written in Java and C. Prominent users of Hadoop are companies like Yahoo, eBay and Facebook just to name a few

2. SAP

The German company Sap, widely known for its ERP solutions, also offer a variety of solutions related to big data The offer is extensive and split up into different categories from analytics to services and

visualizations. One of the interesting applications that SAP offers is Lumira. This application focuses on the visualization of data. It enables the user to create advanced figures and graphs by linking Lumira to the pre-existent database. What makes this interesting is the fact that with this application, SAP stresses the importance of visualization of big data. With Lumira, SAP tried to provide a software tool with a simple user interface to enable staff without extensive training in IT to create visualizations of data that can be used to justify and pre evaluate business decisions.

3. TABLEAU

Tableau is a robust Big Data technology that can be connected to several open-source databases. The server even provides a free public option to create appropriate visualization. This analytics platform consists of various attractive features like sharing options with anybody, moderate speed to enhance extensive operation, integrated with more than 250 applications, and most importantly assists to solve big real-time data analytics issues. It is one of the most powerful, secure, flexible end-to-end real-time data analytics platforms. It generates a series of Tableau product lines— Tableau Prep, Tableau Desktop, Tableau Server, and Tableau Online as well as Tableau Mobile.

4. CASSANDRA

Cassandra is an open-source NoSQL database that transforms multiple sets of real-time data into in-depth analysis. It has linear scalability with proven fault-tolerance on both commodity hardware and cloud infrastructure. Cassandra ensures no data loss while the failed nodes can be replaced efficiently. It has been tested with replay, fuzz, property-based, fault injection as well as multiple performance tests to ensure reliability. It tends to power critical deployments with enhanced performances and scalability in the cloud.

5. OLICK

Click provides transparent raw data integration efficiently with automatically aligned data association. It helps Big Data analysts to detect the potential market trends by integrating embedded and predictive analysis. It supports a full range of real-time data analytics with the Associative Engine and a governed multi-cloud architecture.devices remotely.

6. SPLUNK

Splunk aims to empower IT, DevOps, other teams to transform their multiple sets of real-time data from any source at any time

to detect since the attacker does not change the device's basic functions. In addition, a single device has the potential to infect all smart gadgets in the home.

7. ELASTICSEARCH

Elasticsearch is also an open-source database server utilized for performing full-text search and real-time data analytics with HTTP web interface and Schema-free JSON documents. It is one of the best Big Data technologies due to its reliability and scalability with high speed. It also offers the analysts a smart platform that is highly optimized for language-based searches. It provides rapid results with the implementation of inverted indices for full-text querying, BKD trees, and a column store for real-time data analytics. The scalability can manage Ka jillions of events per second in a 300-node cluster. Elasticsearch is developed alongside the data collection and log -parsing engine Logstash, the analytics and visualization platform kibana, and the collection of lightweight data shippers called Beats. The four products are designed for use as an integrated solution, referred to as the "Elastic Stack

HOW TO USE BIG DATA FOR DECISION MAKING

1. DATA LOGGING

In general, the term Big Data refers to an enormous amount of data that is generated every day by various activities. Among them, we can mention the use of social media, consumption habits and work actions.

By analysing the enormous amount of information recorded during the performance of one of these activities, it is possible to reach interesting conclusions

2. OPTIMIZED PRODUCTION

The implementation of Big Data and Analytics, as well as its integration with ERP and CRM systems, makes it possible for managers to observe, in real time, the most diverse sectors and production stages of their company, facilitating the decision-making.

It is important to emphasize that the use of technology also allows a company's distinguishing features and areas in which it has the best results to be shown more clearly.

3. PERSONALIZED CUSTOMER EXPERIENCE

When it comes to customer service, the benefits of Big Data and Analytics go beyond providing good support. By the interaction between consumers and company, the technology allows their wishes to be assessed and used to provide more personalized services. It is also possible to use the

verified data to observe the emergence of demands in the market and, thus, develop new products to supply them. This makes your customers see it as a cutting-edge enterprise.

4. CLUSTER ANALYSIS

Clustering is a popular unsupervised method and an essential tool for Big Data Analysis. Clustering can be used either as a pre-processing step to reduce data dimensionality before running the learning algorithm, or as a statistical tool to discover useful patterns within a dataset.

5. VISUALISATION ANALYSIS

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets.

6. ASSOCIATION ANALYSIS

Association analysis is the task of finding interesting relationships in large datasets. These interesting relationships can take two forms: frequent item sets or association rules. Frequent item sets are a collection of items that frequently occur together.

7. SENTIMENTAL ANALYSIS

Any Sentimental analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. However, analysis of social media streams is usually restricted to just basic sentiment analysis and count based metrics. Strategy. This is akin to just scratching the surface and missing out on those high value insights that are waiting to be discovered.

6. RESEARCH METHODOLOGIES

HYBRID MODEL

A model may include both descriptive and analytical components. A descriptive model's logical relationships can be examined, and conclusions can be drawn to reason about the system. Nonetheless, logical analysis yields quite different conclusions than a quantitative chemical investigation of system properties.

We first conducted a poll of people utilising an online form creator and data collection service to acquire information regarding people's awareness.

7. PUBLIC SURVEY

We deployed our data gathering utility, often known as a survey bot, to a variety of people and collected information on various facets of their understanding of big data analytics.

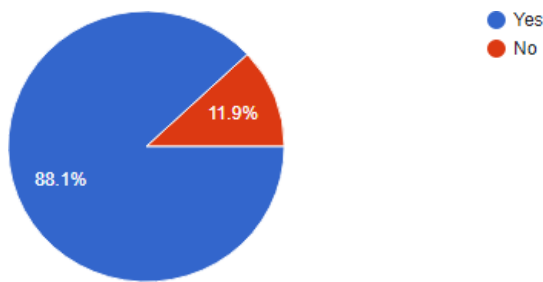
7.1 QUESTIONNAIRE

- Do you have a general understanding of how to access your company's data?
- Do you mostly analyze financial data or do you also look into operational and customer data?
- How vital is it for a company to use Big data for decision making?
- How likely are you to consider using big data in your organization?
- If you could remotely work what are the tools you will be working?
- How safe do you consider big data?
- Do you ever feel big data can help in financial analytics?
- Are you afraid that big data analytics can cause losses to your organization?

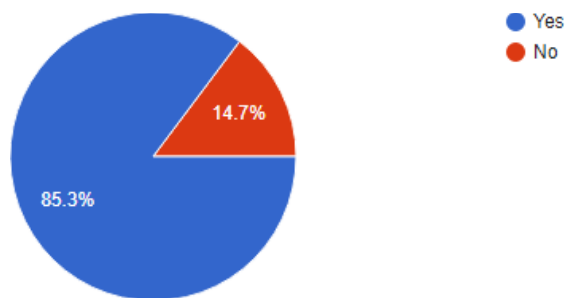
- Would it prohibit you from using big data in your organization for decision making purpose?

7.2 RESULTS

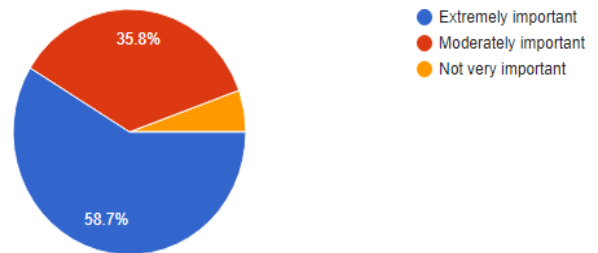
When people were asked if they have idea about how to access company's data then, about 88% were comfortable and the rest were not so comfortable having them around.



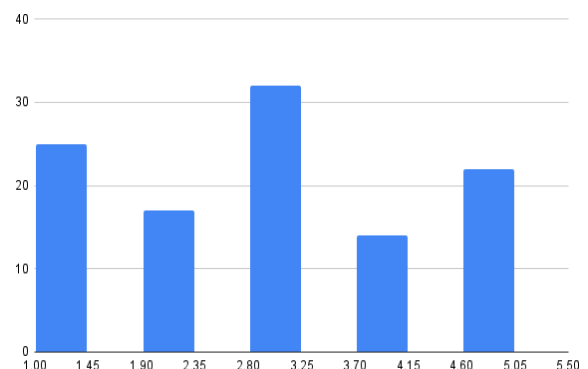
When they were asked if they knew any about financial analytics, or do you look into operational or customer data about 85% were aware of the threats and 15% were not.



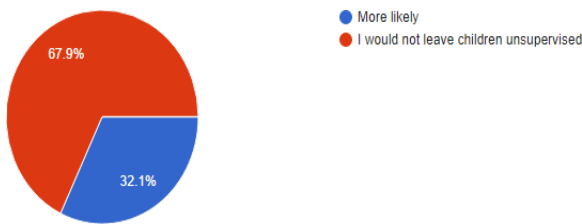
When asked how important or vital is it for a company to use big data, 59% people said it was extremely important for them to be connected to their homes for security reasons, while the rest said it didn't matter to them to stay connected.



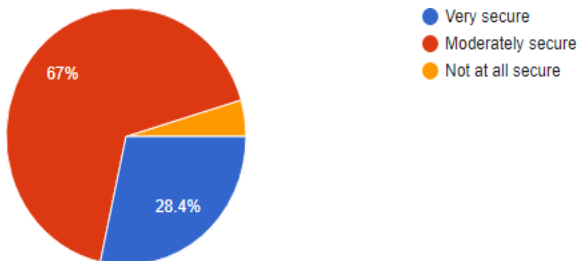
When asked how likely were they to consider using big data in your organization the observed response chart is as follows:



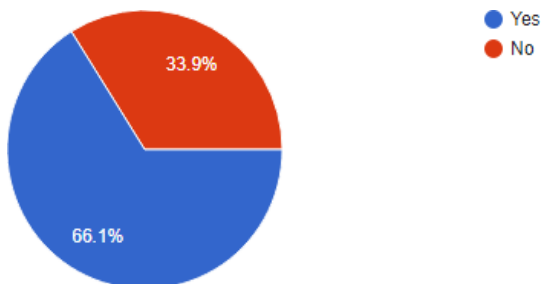
When asked if you remotely work what will be the tools you will be working in, about 68% people said they would not let their children unsupervised while the rest were comfortable if their home were secured by high tech security gadgets.



. When asked how secured do you think big data is, only 28% people were confident that their home network is secured while about 72% of people were not confident about the security of handling big data?

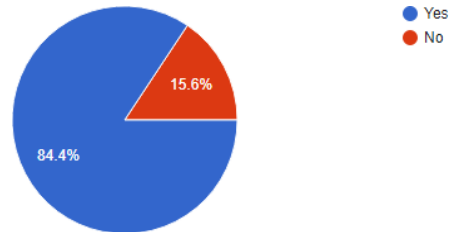


When asked do you think that big data can help in financial analytics, about 66% people said yes.

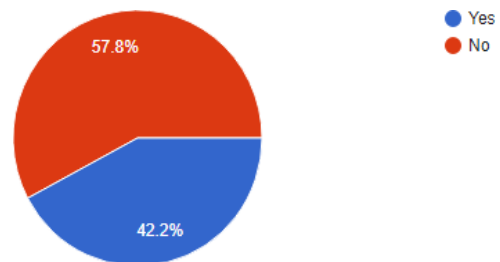


When asked if they were concerned that an big data analytics can cause lose in your organization

about 84% were very concerned while the rest 16% were not.



When asked would financial loss prohibit you using big data in your organization, about 58% people said they would buy smart devices anyway while the rest 42% people said they wouldn't.



8. HYPOTHESIS TESTING

Hypothesis testing is a sort of statistical reasoning that includes analysing data from a sample to derive inferences about a population parameter or probability distribution. First, a hypothesis is created regarding the parameter or distribution. This is known as the null hypothesis, abbreviated as H_0 . After that, an alternative hypothesis (denoted H_a) is

defined, which is the polar opposite of the null hypothesis. Using sample data, the hypothesis-testing technique determines whether or not H_0 may be rejected. The statistical conclusion is that the alternative hypothesis H_a is true if H_0 is rejected.

For this paper,

Null hypothesis (H_0): Smart devices are very secure and can be trusted with our privacy.

Alternative hypothesis (H_a): Smart devices are not secure and cannot be trusted with our privacy.

TEST (STATISTICS)

There are 3 tests available to determine if the null hypothesis is to be rejected or not. They are:

1. Chi-squared test
2. T-student test (T-test)
3. Fisher's Z test.

For this paper, we will be using a 2 tailed T-student test.

A t-test is an inferential statistic that determines if there is a significant difference in the means of two groups that are related in some manner.

- **Level of significance**

(also known as alpha or α). A significance level of 0.05, for example, means there's a 5% probability of discovering a difference when there isn't one. Lower significance levels indicate that more evidence is required to reject the null hypothesis.

- **Level of confidence**

The confidence level indicates the probability that the location of a statistical parameter (such as the arithmetic mean) measured in a sample survey is also true for the entire population.

Sr. No.	Data
1	88.1
2	85.3
3	58.7
4	37.6
5	67.9
6	71.6
7	66.1
8	84.4
9	57.8
Mean (x)	68.611111 11
Standard Deviation (s)	16.202503 24

Level of significance = 0.05 i.e. 5%

Level of confidence = 95%

The chance of rejecting the null hypothesis when it is true is the significance level

A t-score (t-value) is the number of standard deviations away from the t-mean. distribution's.

The formula to find t-score is:

$$t = (x - \mu) / (s / \sqrt{n})$$

where x is the sample mean,

μ is the hypothesized mean,

s is the sample standard deviation, and n is the sample size.

The p-value, also known as the probability value, indicates how probable your data is to have happened under the null hypothesis. Once we know the value of t, we can find the corresponding p-value. If the p-value is less than some alpha level (common choices are .01, .05, and .10) then we can reject the null hypothesis and conclude that smart devices are not secure and cannot be trusted with our privacy.

Calculating t-value:

Step 1: Determine what the null and alternative hypotheses are.

Null hypothesis (H_0): Smart devices are very secure and can be trusted with our privacy.

Alternative hypothesis (H_a): Smart devices are not secure and cannot be trusted with our privacy.

Step 2: Find the test statistic.

In this case, the hypothesized mean value is considered 0.

$$t = (x - \mu) / (s / \sqrt{n}) = (68.61 - 0) / (16.202 / \sqrt{9}) \\ = 12.704$$

t-value = 12.704

Calculating p-value:

Step 3: Calculate the test statistic's p-value.

The t-Distribution table with n-1 degrees of freedom is used to calculate the p-value. In this paper, the sample size is n = 9, so n-1 = 8.

By plugging the observed value in the calculator, it returns a p-value. In this case, the p-value returned is less than 0.00001.

Since this p-value is less than our chosen alpha level of 0.05, we can reject the null hypothesis. Thus, we have sufficient evidence to say that smart devices are not secure and cannot be trusted with our privacy.

9. FINDINGS

1. Big data typically refers to large volumes of information systems which can be analyzed and create important insights to help businesses make better decisions.

2. For instance, cognitive computing analyses data to streamline in-office processes and operations in both

the public and private sectors. Big data applications can also be used to improve the overall management of a business.

3. Big data includes a mix of structured, semi-structured and unstructured real-time data, consisting of OLAP, ETL and information. Business firms and professionals extract useful information from big data to help inform business decisions

4. It is not only used by companies to make real-time decisions for improving business results but big data also adds a competitive advantage over different firms in the marketplace. Furthermore, data analytics plays a major role in creating a framework for managing data in decision making, thereby assisting business owners to make better decisions for their firms.

5. While handling large volumes of data may sound very technical, becoming proficient in the practice will increase the efficiency of your business. The key to drawing real value from data analytics is to determine how successful your data-driven decisions are.

6. In order to become a more data-driven professional who can turn data into insights,

10. CONCLUSION

Data analysis is important in business as it helps us understand problems faced by

an organisation and effectively use data to solve these issues. Data consists of facts and figures, and data analysis organizes, interprets, structures and presents that data into useful information. This context can then be further used by decision-makers to enhancing productivity and increase business profits. Following is a list of the advantages of big data technologies for business. Data can effectively improve overall management and eventually change how businesses operate for the better. With the continuous improvement of data technology, the demand for educated business professionals who specialize in data analytics is increasing rapidly.

11. BIBLOGRAPHY

1. <https://www.techopedia.com/definition/31463/smart-device>
2. <https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT>
3. <https://builtin.com/iot-internet-things/smart-device>
4. <https://towardsdatascience.com/what-is-a-smart-device-the-key-concept-of-the-internet-of-things-52da69f6f91b>
5. <https://home.howstuffworks.com/smart-home.htm#pt1>

6.

<https://dl.acm.org/doi/fullHtml/10.1145/3290605.3300498>

7.

<http://smarthomes-technology.blogspot.com/p/conclusion-and-additional-work.html>

8.

<https://www.kaspersky.co.in/resource-center/threats/how-safe-is-your-smart-home>

9.

<https://us.norton.com/internetsecurity-iot-smart-home-security-core.html>