

# IDEATION PHASE

## LITERATURE SURVEY

Date	29 August 2022
Team ID	PNT2022TMID07696
Project Name	Estimate The Crop Yield Using Data Analytics

### LITERATURE SURVEY FOR DATA ANALYTICS:

#### Abstract:

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques.

Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Keywords: big data, data mining, analytics, decision making.

#### Introduction:

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages.

Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc.

Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet. With the increase in storage capabilities and methods of data collection, huge amounts of data have become easily available.

Every second, more and more data is being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data. The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted.

## 1 A Literature Review :

The contribution of this paper is to provide an analysis of the available literature on big data analytics. Accordingly, some of the various big data tools, methods, and technologies can be applied are discussed, and their applications and opportunities provided in several decision domains are portrayed.

The literature was selected based on its novelty and discussion of important topics related to big data, in order to serve the purpose of our research. The publication years range from 2008-2013, with most of the literature focusing on big data ranging from 2011-2013. This is due to big data being a recently focused upon topic. Furthermore, our corpus mostly includes research from some of the top journals, conferences, and white papers by leading corporations in the industry.

## 2 Big Data Analytics

The term “Big Data” has recently been applied to datasets that grow so large that they become awkward to work with using traditional database management systems. They are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process the data within a tolerable elapsed time [12]. Big data sizes are constantly increasing, currently ranging from a few dozen tera-bytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before [17].

Hence, big data analytics is where advanced analytic techniques are applied on big data sets.

Therefore the successive section will elaborate the big data analytics tools and methods, in particular, starting with the big data storage and management, then moving on to the big data analytic processing. It then concludes with some of the various big data analyses which have grown in usage with big data.

### 2.1 Characteristics of Big Data

Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures, analytics, and tools in order to enable insights that unlock new sources of business value.

Three main features characterize big data: volume, variety, and velocity, or the three V's. The volume of the data is its size, and 216 N.

Elgendy and A. Elragal how enormous it is. Velocity refers to the rate with which data is changing, or how often it is created. Finally, variety includes the different formats and types of data, as well as the different kinds of uses and ways of analyzing the data [9].

Data volume is the primary attribute of big data. Big data can be quantified by size in TBs or PBs, as well as even the number of records, transactions, tables, or files.

## 2.2 Big Data Analytics Tools and Methods

With the evolution of technology and the increased multitudes of data flowing in and out of organizations daily, there has become a need for faster and more efficient ways of analyzing such data. Having piles of data on hand is no longer enough to make efficient decisions at the right time.

**Big Data Storage and Management** One of the first things organizations have to manage when dealing with big data, is where and how this data will be stored once it is acquired. The traditional methods of structured data storage and retrieval include relational databases, data marts, and data warehouses.

The data is uploaded to the storage from operational data stores using Extract, Transform, Load (ETL), or Extract, Load, Transform (ELT), tools which extract the data from outside sources, transform the data to fit operational needs, and finally load the data into the database or data warehouse.

Thus, the data is cleaned, transformed, and catalogued before being made available for data mining and online analytical functions [3]. However, the big data environment calls for Magnetic, Agile, Deep (MAD) analysis skills, which differ from the aspects of a traditional Enterprise Data Warehouse (EDW) environment.

# 3 Big Data Analytics and Decision Making :

From the decision maker's perspective, the significance of big data lies in its ability to provide information and knowledge of value, upon which to base decisions. The managerial decision making process has been an important and thoroughly covered topic in research throughout the years.

Big data is becoming an increasingly important asset for decision makers. Large volumes of highly detailed data from various sources such as scanners, mobile phones, loyalty cards, the web, and social media platforms provide the opportunity to deliver significant benefits to organizations. This is possible only if the data is properly analyzed to reveal valuable insights, allowing for decision makers to capitalize upon the resulting opportunities from the wealth of historic and real-time data generated through supply chains, production processes, customer behaviors, etc. [4].

However, the need for analyzing external data, such as customer markets and supply chains, has arisen, and the use of big data can provide cumulative value and knowledge. With the increasing sizes and types of un-structured data on hand, it becomes necessary to make more informed decisions based on drawing meaningful inferences from the data [7].

Accordingly, [8] developed the B-DAD framework which maps big data tools and techniques, into the decision making process [8].

Such a framework is intended to enhance the quality of the decision making process in regards to dealing with big data.

The first phase of the decision making process is the intelligence phase, where data which can be used to identify problems and opportunities is collected from internal and external data sources.

In this phase, the sources of big data need to be identified, 222 N. Elgendy and A. Elragal and the data needs to be gathered from different sources, processed, stored, and mi-grated to the end user.

Such big data needs to be treated accordingly, so after the data sources and types of data required for the analysis are defined, the chosen data is ac-quired and stored in any of the big data storage and management tools previously discussed After the big data is acquired and stored, it is then organized, prepared, and processed, This is achieved across a high-speed network using ETL/ELT or big data processing tools, which have been covered in the previous sections.

## 3.1 Customer Intelligence

Big data analytics holds much potential for customer intelligence, and can highly benefit industries such as retail, banking, and telecommunications. Big data can create transparency, and make relevant data more easily accessible to stakeholders in a time-ly manner [14].

Big data analytics can provide organizations with the ability to profile and segment customers based on different socioeconomic characteristics, as well as increase levels of customer satisfaction and retention [4].

This can allow them to make more informed marketing decisions, and market to different segments based on their preferences along with the recognition of sales and marketing opportunities [17].

Additionally, using SNAs to monitor customer sentiments towards brands, and identify influential individuals, can help organizations react to trends and perform direct marketing. Big data analytics can also enable the construction of predictive models for customer behavior and purchase patterns, therefore raising overall profitability [4].

Even organizations which have used segmentation for many years are beginning to deploy more sophisticated big data techniques, such as real-time micro-segmentation of customers, in order to target promotions and advertising [14].

Consequently, big data analytics can benefit organizations by enabling better targeted social influencer marketing, defining and predicting trends from market sentiments, as well as analyzing and understanding churn and other customer behaviors [17].

## 3.2 Supply Chain and Performance Management

As for supply chain management, big data analytics can be used to forecast demand changes, and accordingly match their supply.

This can increasingly benefit the manufacturing, retail, as well as transport and logistics industries. By analyzing stock utilization and geospatial data on deliveries, organizations can automate replenishment decisions, which will reduce lead times and minimize costs and delays, as well as process interruptions. Additionally, decisions on changing suppliers, based on quality or price competitiveness, can be taken by analyzing supplier data to monitor performance.

Furthermore, alternate pricing scenarios can be run instantly, which can enable a reduction in inventories and an increase in profit margins [4].

Accordingly, big data can lead to the identification of the root causes of cost, and provide for better planning and forecasting [17].

Another area where big data analytics can be of value is performance management, where the governmental and healthcare industries can easily benefit.

This can allow departments to link their strategic objectives with the service or user outcomes, thus leading to increased efficiencies.

### 3.3 Quality Management and Improvement

Especially for the manufacturing, energy and utilities, and telecommunications industries, big data can be used for quality management, in order to increase profitability and reduce costs by improving the quality of goods and services provided.

For example, in the manufacturing process, predictive analytics on big data can be used to minimize the performance variability, as well as prevent quality issues by providing early warning alerts.

This can reduce scrap rates, and decrease the time to market, since identifying any disruptions to the production process before they occur can save 224 N. Elgendy and A. Elragal significant expenditures [4].

Additionally, big data analytics can result in manufacturing lead improvements [17].

Furthermore, realtime data analyses and monitoring of machine logs can enable managers to make swifter decisions for quality management.

Also, big data analytics can allow for the real-time monitoring of network demand, in addition to the forecasting of bandwidth in response to customer behavior.

Additionally, the quality of citizens' lives can be improved through the utilization of big data. For healthcare, sensors can be used in hospitals and homes to provide the continuous monitoring of patients, and perform real-time analyses on the patient data streaming in.



This can be used to alert individuals and their health care providers if any health anomalies are detected in the analysis, requiring the patient to seek medical help [22].

Patients can also be monitored remotely to analyze their adherence to their prescriptions, and improve drug and treatment options [14].

Moreover, by analyzing information from distributed sensors on handheld devices, roads, and vehicles, which provide real-time traffic information, transportation can be transformed and improved.

Traffic jams can be predicted and prevented, and drivers can operate more safely and with less disruption to the traffic flow. Such a new type of traffic ecosystem, with “intelligent” connected cars, can potentially renovate transportation and how roadways are used [22].

Accordingly, big data applications can provide smart routing, according to real-time traffic information based on personal location data. Furthermore, such applications can automatically call for help when trouble is detected by the sensors, and inform users about accidents, scheduled road-work, and congested areas in real-time [14]. Furthermore, big data can be used for better understanding changes in the location, frequency, and intensity of weather and climate.

This can benefit citizens and businesses that rely upon weather, such as farmers, as well as tourism and transportation companies.

Also, with new sensors and analysis techniques for developing long term climate models and nearer weather forecasts, weather related natural disasters can be predicted, and preventive or adaptive measures can be taken beforehand [22].

## 3.4 Risk Management and Fraud Detection

Industries such as investment or retail banking, as well as insurance, can benefit from big data analytics in the area of risk management.

Since the evaluation and bearing of risk is a critical aspect for the financial services sector, big data analytics can help in selecting investments by analyzing the likelihood of gains against the likelihood of losses.

In addition, customer intelligence can be used to model normal customer behavior, and detect suspicious or divergent activities through the accurate flagging of outlier occurrences.

Furthermore, providing systems with big data about prevailing fraud patterns can allow these systems to learn the new types of frauds and act accordingly, as the fraudsters adapt to the old systems designed to detect them. Also, SNAs can be used to identify the networks of collaborating fraudsters, as well as discover evidence of fraudulent insurance or benefits claims, which will lead to less fraudulent activity going undiscovered [4].

Thus, big data tools, techniques, and governance processes can increase the prevention and recovery of fraudulent transactions by dramatically increasing the speed of identification and detection of compliance patterns within all available data sets [22].

## 4 Conclusion

In this research, we have examined the innovative topic of big data, which has recently gained lots of interest due to its perceived unprecedented opportunities and benefits.

In the information era we are currently living in, voluminous varieties of high velocity data are being produced daily, and within them lay intrinsic details and patterns of hidden knowledge which should be extracted and utilized.

Hence, big data analytics can be applied to leverage business change and enhance decision making, by applying advanced analytic techniques on big data, and revealing hidden insights and valuable knowledge. Accordingly, the literature was reviewed in order to provide an analysis of the big data analytics concepts which are being researched, as well as their importance to decision making. Consequently, big data was discussed, as well as its characteristics and importance.

With all the problems faced with traditional data management, big data exponentially increases these difficulties due to additional volumes, velocities, and varieties of data and sources which have to be dealt with. Therefore, future research can focus on providing a roadmap or framework for big data management which can encompass the previously stated difficulties.

We believe that big data analytics is of great significance in this era of data over-flow, and can provide unforeseen insights and benefits to decision makers in various areas. If properly exploited and applied, big data analytics has the potential to provide a basis for advancements, on the scientific, technological, and humanitarian levels.

## References

1. Adams, M.N.: Perspectives on Data Mining.
2. International Journal of Market Research 52(1), 11–19 (2010)
3. 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)
4. Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)
5. Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)

6. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492 (2009)
7. Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data:
8. The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011)
9. Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making.