RESEARCH ARTICLE OPEN ACCESS

Big Data Handshakes Drones & Sensors in Precision Agriculture

Manish Kumar [1], Dr. Prof. Deva Prakash [2], Seema Sinha [3]
[1] Research Scholar, Department of Mathematics, Magadh University, Bodhgaya
[2] Associate Professor, Head of Department, Department of Mathematics, S.M.D. College, Punpun
[3] Research Scholar, Department of Mathematics, Magadh University, Bodhgaya - Bihar

ABSTRACT
Agriculture is the lifeblood of India's rural population. It is also necessary for human survival, and it has evolved over time to meet the needs of an ever-increasing human population. They rely on agriculture to support themselves. Agriculture [1] has a long history in India, dating back to the Rigveda. [2] India now ranks second in the world for agricultural output. Agriculture and related industries such as forestry and fisheries accounted for 13.7 percent of GDP in 2013,[3] and employed about half of the workforce. [4] [5] The share of agriculture in GDP increased to 20.2 per cent in 2020-21 from 18.4 per cent in 2019-20. [5] The last time the contribution of the agriculture sector in GDP was at 20 per cent was in 2003-04. Agriculture, however, is India's most populous economic sector and plays a key role in the country's entire socioeconomic fabric. With India's broad-based economic expansion, agriculture's economic contribution to GDP is rapidly shrinking. India's agricultural exports were $50.21 billion in 2021-22 [5], making it the world's ninth largest agricultural exporter in 2019. The majority of its agricultural exports are destined for developing and least developed countries. [5] Indian agricultural/horticultural and processed goods are exported to over 100 countries, primarily in the Middle East, Southeast Asia, the SAARC countries, the European Union, and the United States. [6] More than 60% of the crop is still reliant on monsoon rains. Recent advancements in information technology for agriculture have sparked an interest in crop yield prediction research. This study examines the consequences of old farming techniques and discusses how modern computer technology can be used to boost agricultural commodity yields. It also recognises Big Data's important computational and diagnostic capabilities in real-time processing of large volumes of transactional data. Another goal of this article is to offer agricultural modifications and to stimulate conversation about how innovations in Big Data may nurture to improve the rural agriculture system.
Keywords: - Big Data, Big Data Analytics, Agriculture Drones, Electronic Sensors, Rural agricultural systems, Precision Agriculture, Electronic farm records, Hadoop framework, Spark

I. INTRODUCTION

Knowing what to grow and when to plant it has always been as much an art as a science for farmers. The relevance of a country's Gross Domestic Product (GDP) and its capacity to support feeding and equipping its own inhabitants, in addition to exporting raw materials and completed goods, is critical for a healthy and prosperous economy.

By no means can the percentage of GDP growth based on year-over-year numbers be used to gauge development. Again, industrialization cannot be called the trailblazer in an integrated economy because agriculture and related sectors (Subsectors include agriculture, livestock, forestry, and fishing) will always be a big contributor to any economy in the globe.

The total share of agriculture and allied sectors in terms of percentage of GDP was 20.2 percent during 2020-21.[9]

Agriculture, along with its related industries, is India's most important source of income. Agriculture continues to be the primary source of income for 70% of rural households, with 82 percent of farmers being small and marginal. Total food grain output was predicted to be 275 million tons in 2017-18. (MT). India is the world's largest producer of pulses (25 percent of global output), consumer (27 percent of global consumption), and importer (14 percent). India's annual milk output was 165 MT in 2017-18, making it the world's largest producer of milk, jute, and pulses, and with 190 million cattle in 2012. [10] It is the world's second-largest producer of rice, wheat, sugarcane, cotton, and groundnuts, as well as the second-largest producer of fruits and vegetables, accounting for 10.9 percent of global fruit and vegetable output and 8.6 percent of global fruit and vegetable production, respectively.

However, India continues to face several developing issues. Agriculture's proportion to GDP has consistently fallen from 1951 to 2011, as the Indian economy has diversified and risen. Despite reaching food self-sufficiency in production, India still has a quarter of the world's hungry people and over 190 million people who are undernourished.

The following are some of the key factors that improve agricultural system performance and productivity:

The first and most important consideration is

1. How to store enormous amounts of unstructured data.
2. How can data access, storage, and analysis be accelerated?
3. How can large amounts of data be analyzed in real time?
4. How might the power of actionable analytics help you respond to situations quickly?

Most socially aware individuals I encounter ask me, "How can our big data assist agriculture or a poor farmer?"

We are an agricultural country, with 60% of the population dependent on agriculture directly or indirectly. When I discuss the potential of what big data can do for agriculture and farmers, they are enthusiastic at first, but quickly become gloomy, asking how long it will take. Who will be in charge of gathering such a wide range of data? How many farmers would commit suicide before the bomb exploded? I'm blank, but I do share an optimistic perspective.

A. Precision Agriculture

Precision agriculture (PA) refers to the application of exact and right amounts of inputs such as water, fertilizers, insecticides, and other chemicals to crops at the appropriate times in order to increase productivity and maximize harvests.

PA originated in the United States and Europe. Because the farms in these nations are typically large (> 100 hectares), GPS technology is often employed in PA. GPS enables for exact mapping of farms, which, when combined with proper software, informs farmers about the state of their crops and which parts of the farm require inputs like water or fertilizer.

Precision agriculture is also distinguished in western nations by the use of heavy farm machinery (average power of the machinery is 100-200 kW) for all farm and field operations such as sowing, harvesting, weeding, baling, and so on. This equipment runs on fossil fuels and consumes more than 60% of the total energy used in agriculture. Furthermore, heavy farm gear compacts the soil, rendering it unproductive.

On the other hand, PA for small farms can employ tiny farm gear and robots that do not compact the soil and can run on sustainable fuels such as bio-oil, compressed biogas, and electricity generated on farms from agricultural leftovers.

Precision agriculture for small farms may feature sub-surface drip irrigation for accurate water and fertiliser delivery, as well as robots for no-till planting, weed eradication, harvesting, and other tasks. Small farms in the United States and Europe are already using some of these robots.

Drones have also been adopted in Japan and the United States for mapping farms, diagnosing illnesses, and other purposes. Because most robotic devices and drones are tiny, they are ideal for small farms. As a result, India's tiny farms are perfect for precision agriculture on a big scale.

II. INDIA’S CURRENT AGRICULTURAL SYSTEM

In India, farming systems are deliberately used in regions where they are most appropriate. Subsistence farming, organic farming, and industrial farming are the farming methods that contribute considerably to India's domestic GDP. [1] The forms of farming used in different parts of India vary; some rely on horticulture, ley farming, agroforestry, and other methods. [1] Because of India's geographical position, different sections of the country have varying climates, which impact agricultural output differently. For significant crop yields, India relies heavily on its monsoon cycle.

Agriculture in India has a long history, dating back at least ten thousand years. Currently, the country ranks second in the world in agricultural production. Agriculture and other Industry counted for more than 16 percent of India's GDP in 2007. Despite the fact that agriculture's contribution to the country's GDP has been steadily declining, agriculture remains India's largest industry and plays a critical part in the country's socioeconomic development. Wheat, rice, cotton, sugarcane, silk, groundnuts, and a variety of other crops are all produced in India.

It is also the second-largest vegetable and fruit harvester, accounting for 8.6% and 10.9 percent of total production, respectively. Mangoes, papayas, sapotas, and bananas are among India's most popular fruits. India also boasts the world's largest cattle population, with 281 million animals. With 175 million cattle, the country was the second biggest in the world in 2008.[3]

III. BIG DATA IN GENERAL

A. Introduction of Big Data

Historically, the term Big Data is quite vague and ill defined. It is not a precise term and does not carry a particular meaning other than the notion of its size. The word “big” is too generic; the question how “big” is big and how “small” is small [11] is relative to time, space, and circumstance. From an evolutionary perspective, the size of “Big Data” is always evolving. If we use the current global Internet traffic capacity [12] as a measuring stick, the meaning of Big Data volume would lie between the terabyte (TB or 1012 or 240) and zettabyte (ZB or 1021 or 270) range. Based on the historical data traffic growth rate, Cisco claimed that humans have entered the ZB era in 2015 [12].

B. The origin of Big Data

Several research on historical perspectives and developments in the BDA region have been done. Gil Press [13], based on Rider's work [14], offered a brief history of Big Data beginning in 1944. Between 1944 and 2012, he covered 68 years of Big Data evolution and presented 32 Big Data-related events in modern data science history. The narrow line between data expansion and Big Data has become blurred, as Press pointed out in his piece. The rate of data expansion is sometimes referred to as a 0"information explosion," despite the fact that the terms "data" and "information" are frequently used interchangeably. The analysis by Press is fairly extensive, including BDA occurrences up until December 2013.

In contrast to Press' analysis, Frank Ohlhorst [15] dates the birth of Big Data to 1880, the year of the tenth US census. The main difficulty in the nineteenth century was a statistical problem: how to survey and document 50 million people in North America. Despite the fact that Big Data may include statistical calculations, the two phrases have different meanings nowadays.

Fig 1: A short history of big data.

Similarly, Winshuttle [16] believes Big Data began in the nineteenth century. Winshuttle claims that data sets that are so massive and complicated that they are beyond typical process and management capacity can be classified as Big Data. In contrast to Press' review, Winshuttle's study focuses on enterprise resource planning and cloud infrastructure implementation. In contrast to Press' review, Winshuttle's study focuses on enterprise resource planning and cloud infrastructure implementation.

Visualizing.org gave another historical review. [17] It concentrated on the timing for implementing BDA. The events surrounding the Big Data push by major Internet and IT businesses, such as Google, YouTube, Yahoo, Facebook, Twitter, and Apple, have largely shaped its historical depiction. It stressed Hadoop's substantial effect on BDA's history. It mostly focused on Hadoop's crucial role in the BDA.

Douglas Laney's 3Vs description has been widely accepted as the "common" features of Big Data, according to the history of the Big Data timeline [18], although he stopped short of attaching these traits to the phrase "Big Data."

IV. DIFFERENT ATTRIBUTES OF BIG DATA DEFINITIONS

A. Gartner — 3Vs definition

Many properties have been added to Big Data since 1997. Three of these characteristics are the most well-known and have been frequently acknowledged and implemented. The first is what is known as Gartner's interpretation, or 3Vs; the name dates back to February 2001. Douglas Laney [18] predicted it in a white paper issued by Meta Group, which was later bought by Gartner in 2004. Data has risen in three dimensions as a result of the surge in e-commerce operations, according to Douglas:

1. Volume, which refers to both the incoming data stream and the total amount of data.
2. Velocity, which is the rate at which data is utilised to support and created by interactions.
3. Variety, which denotes a wide range of incompatible and inconsistent situations.

B. IBM — 4Vs definition

On top of Douglas Laney's 3Vs nomenclature, which is known as the 4Vs of Big Data, IBM added another characteristic or "V" for "Veracity." Each "V" is defined as follows [19],[20]:

1. Data scale is represented by volume.
2. Analysis of flowing data is referred to as velocity.
3. Variety denotes several types of data.
4. Data uncertainty is implied by veracity.

The extra "V" or veracity component was added by Zikopoulos et al. "in response to the quality and source challenges our clients were confronting with their Big Data projects" [21].

C. Microsoft - 6Vs definition

Microsoft expanded Douglas Laney's 3Vs qualities to 6Vs [22], adding variability, veracity, and visibility in order to maximise commercial value:

1. Data scale is represented by volume.
2. Analysis of flowing data is referred to as velocity.
3. Variety denotes several types of data.
4. Veracity is concerned with the reliability of data sources.
5. The complexity of the data collection is referred to as variability. It refers to the number of variables in data sets, as opposed to "variety" (or a distinct data format).
6. Visibility stresses the importance of having a complete view of facts while making informed decisions.

D. More Vs for big data

In 2013, Yuri Demchenko [23] presented a 5 Vs' Big Data description. Along with the IBM 4Vs definition, he introduced the value dimension (see Fig. 2). Since Douglas Laney's publication of 3Vs in 2001, there have been more "Vs," up to 11 [24].

All of these definitions, such as 3Vs, 4Vs, 5Vs, and even 11Vs, are attempting to express the feature of data. The majority of them are data-oriented definitions, but they don't adequately describe Big Data in connection to BDA's essence. We must first define data in order to comprehend the key meaning.

Within the cosmos, data is everything. This indicates that data is limited by current technical capabilities. There are no data boundaries or limitations if the technical capability is permitted. The question is why we should even try to capture it. Clearly, the major goal for gathering data is not to discover a better solution for our research or business problem, which is to look for actionable information, but to find a better answer for our research or business problem. Pure data-driven analysis may provide minimal benefit to a decision maker; in certain cases, it may simply increase the burden of BDA expenditures or resources. Perhaps this is why Harper thinks Big Data is so difficult. [25]

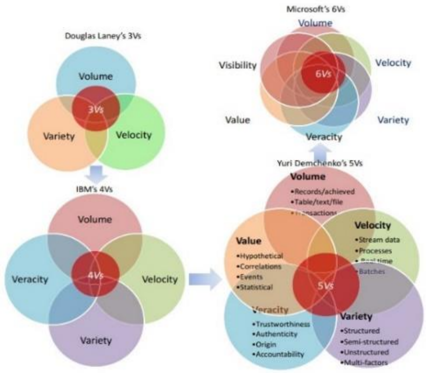

Fig 2: From 3Vs, 4Vs, 5Vs, and 6Vs big data definition.

V. BIG DATA EXAMPLES

A. Fraud detection

Fraud detection is one of the most appealing Big Data application examples for firms whose operations entail any form of claims or transaction processing. Detecting fraud on the fly has always been a difficult task. The majority of the time, fraud is detected after the fact, when the damage has already been done and all that remains is to minimize the damages and alter procedures to prevent it from happening again. Big Data solutions that can monitor claims and transactions in real time, discovering large-scale trends across multiple transactions or detecting unusual behaviour from a single user, can alter the game when it comes to fraud detection.[26]

B. IT log analytics

Solutions and departments produce a massive number of logs and trace data. Much of this data must go unexplored in the absence of a Big Data solution: enterprises just lack the people or resources to sift through all of this data by hand, much alone in real time. Those logs and trace data, on the other hand, may be put to good use with a Big Data solution in place. IT log analytics is the most widely applicable of these Big Data application examples. Any company with a big IT department will benefit from the capacity to swiftly discover large-scale trends to aid in issue diagnosis and prevention. Similarly, any company with a big IT staff will value the opportunity to track incremental improvements.[26]

C. Call center analytics

Now we'll look at instances of customer-facing Big Data applications, with call center analytics being particularly useful. Without a Big Data solution, most of the knowledge that a call center may give would be ignored or found too late. Not just by interpreting time/quality resolution indicators, but also by recording and analyzing call content, Big Data solutions may assist discover reoccurring problems or customer and staff behaviour patterns on the fly.[26]

D. Social media analysis

Analysis of social media activity is one of the most essential customer-facing Big Data application examples we might cover. These days, everyone and their mothers is on social media, whether it's "liking" corporate pages on Facebook or tweeting product complaints on Twitter. IBM's Cognos Consumer Insights, a point solution based on IBM's Big Insights Big Data platform, is a Big Data solution designed to capture and analyze social media activity. Social media may give real-time information on how customers react to items and campaigns. Companies can benefit from these insights.[26]

For the best results, they modify their price, marketing, and campaign placement on the fly.

E. Financial Services

Using Bigdata analytics, forecast client bank-related actions such as insurance and debit card usage. In recent years, the financial services industry has seen significant changes.

Customers want their banks to provide a more personalized/private data service.

Bigdata has the potential to transform the way businesses operate by capturing the optimal value in consumer data and re-modeling market interactions. Bigdata can include internal and external data from a variety of sources [8].

F. Airlines and Tracking Companies

Bigdata is used to track fuel use and airline habits in order to improve competence and efficiency. Bigdata is now defined as a collection of data with a large volume, velocity, and diversity. It has the potential to provide economic value to operations, decision-making, risk management, and customer service. Airlines are terrific at gathering data, but they aren't necessarily good at exploiting it. Even as airlines acquire more and more data, massive amounts of data are stored, processed, and then dropped in the airline sector. Terabytes of client data are floating about in the system at any given time. Bigdata Analytics is in charge of these projects.[8]

G. Healthcare Providers

Drug efficiency is measured using healthcare practitioners. The healthcare providers are distinguished by specific patient results and experiences, as well as patient and medical personnel referrals. It also helps with cost management and more flexible patient monitoring. To increase care quality while lowering costs. [8]

H. Telecommunications

In order to investigate user behaviour and demand trends. Telecom firms are quite common nowadays. They have a lot of data, and they need to go through it and analyze it properly, both organized and unstructured data. A multinational mobile communications company understands customer location and travel habits. To identify the effects of fraud while minimizing the volume of assistance by utilizing Bigdata. The consumption of data in the telecom sector will increase at an annual pace of data.[8]

Finally, Bigdata refers to the ability to

- identify, aggregate, and handle many data sources.
- Create complex analytics models to predict and process results.
- Analyze organizational data to make better decisions.

VI. OBJECTIVES OF RESEARCH

Investments in agriculture, according to Bill Gates, co-chair of the Bill & Melinda Gates Foundation, are the finest weapons against hunger and poverty.

The goal of this paper is to boost agricultural production to meet the growing global population.

- To lower the costs of agricultural cultivation.
- To identify the issues that farmers face in India.
- To expand the international market for Indian agricultural goods.
- To determine how the seasons of the climate change.
- To fine-tune weather forecasts and historical data for agricultural choices.

VII. REVIEW OF LITERATURE ON BIG DATA IN AGRICULTURE

A. G. RajeshKumar et al., (2015): Agriculture has an essential role in the economic growth and food security of agro-based countries. Crop selection is the most important aspect of agricultural planning. In the agricultural business, several characteristics are employed, including productivity, market rate, market farm, government regulations, and policies.

Crop production rate, weather factors, soil conditions, and crop categorization are all calculated using machine learning approaches.

Crop selection is the approach covered in this study. This strategy is mostly used to correct the crop selection problem and boost the country's maximum economic development. This also maximizes the agricultural data's net yield rate.

Techniques Involved

SVM (Support Vector Machine): The support vector machine is another name for crop yield forecasting. Using the kernel function, derive the non-linear function.

Gradient Boosted Decision Tree (GBDT): This is an additive decision tree approach in which a series of decision trees are combined to form a forest that may be used as a predictive model.

Regularized Greedy Forest (RGF) is a decision tree technique that is also additive.

The RGF generates a globally optimal decision tree, similar to the GBDT.

These algorithms have the following advantages:

(i) Linear functions are employed to overcome issues in SVM. In GBDT, the base learner may be swapped to another wrapper.

(ii) In RGF, it works quicker and more precisely.

These strategies have the following drawbacks:

(i) Boosting wrapper treats decision tree as a Black box.

(ii) GBDT allows for tree optimization rather than forest optimization. [27]

B. Rupika Yadhav et al., (2015): This research examines a collection of traditional farming applications and corrects how to enhance the productivity of agricultural elements by utilizing current technology. The government's role in improving rural agricultural systems is also discussed in this study. HDFS, Map Reduce, Hadoop, STORM, and other e-agriculture services are employed here. Irrigation, market infrastructure, and delivery operations are all issues that have arisen in Indian agriculture. In Indian agriculture, the best delivery techniques are not accessible. The government tried numerous ways to handle data in the agricultural business but they failed, therefore the Bigdata idea was introduced to gather, store, and analyze data in agriculture. The goal of this study is to apply many farming methods, including the utilization of such as use of pesticides and use of fertilizers. [28]

Techniques Involved: Soil temperature maps and data, precipitation and data, electrical conductivity and data, moisture content data, and other information are stored in Electronic Farm Records [EFR]. Finally, the major function of Bigdata professionals is to uncover the linkages, as well as trends to increase agricultural yield through ideal diagnostic systems.

The advantages of this approach include

(i) soil and crop sensing, which allows agriculture to become more efficient.

(ii) It is less costly and produces higher-quality results than the prior method.

(iii) This paper's implementation might reveal information hidden in agricultural databases.

C. Hemlata Channe et al: In this study, we will create an Agro cloud to store information on farmers, cyclic soil qualities of farmlands, vendors, marketing, and e-governance in the Agro sector, as well as current environmental circumstances. Utilize the Beagle black bone, which may be interfaced to soil and environmental features, to maximize agricultural yields, oil properties, and current environmental conditions.

The disadvantage is that Bigdata approaches necessitate new frameworks for efficiently storing data. [29]

D. Duncan Waga et al., (2014): Environmental aspects such as winds, temperature, and rainfall are discussed in this work. It also makes use of a unique cloud computing analytical tool to extract essential information that farmers may utilize for critical and profitable agriculture. Ubuntu and eucalyptus are used to create a private cloud. Apache Flume is used to gather, aggregate, and disseminate data into HDFS in a distributed, flexible, and efficient manner. It has a flexible and straightforward architecture. [30]

The benefit is that it is adaptable, dedicated, and dispersed.

The drawbacks are:

There are two sorts of challenges in Big Data analytics employing cloud storage:

(i) Capacity and performance, and scaling and capacity.

(ii) Hadoop has been integrated into the new IT platform.

E. Ling Zhang et al., (2009): The Agriculture Information Degree (AID) model is introduced in this work, which is based on self-adapting regression and implemented using the agriculture information degree index system. It employs a number of classifiers and also ensures the correctness and efficiency of agricultural data.

The RBF kernel functions are used in the self-adapting tends vector regression procedure, and there is no sensitivity coefficient £.

The benefits include I avoiding parameters associated with complex certification processes.

(ii) To improve the system's accuracy and performance.

(iii) To overcome the flaws that exist in conventional methods and to increase generalization capabilities. [31]

VIII. DIFFERENT TOOLS AND TECHNIQUES USED FOR BIG DATA ANALYTICS

The tools that can be used to analyze agricultural data at the moment are described below.

A. Weka: Weka (Waikato Environment for Knowledge Analysis) is a popular Java-based machine learning software package created at New Zealand's University of Waikato. The Weka workbench includes a set of visualization tools and algorithms for data analysis and predictive modelling, as well as graphical user interfaces that make this capability accessible. Weka can do data preparation, clustering, classification, regression, visualization, and feature selection, among other data mining activities.

Weka is a set of machine learning algorithms that may be used to solve real-world data mining challenges. It's written in Java and works on nearly every platform. The algorithms may be applied on a dataset directly or invoked from our own Java code. [32]

Weka has the following advantages:
1. It is available for free under the GNU General Public License.
2. Portability, as it is written entirely in the Java programming language and so operates on virtually any current computing platform.
3. An extensive library of data preparation and modelling techniques
4. Its graphical user interfaces make it simple to utilize.

B. MATLAB: MathWorks created the MATLAB programming language. It began as a matrix programming language with rudimentary linear algebra programming. It may be used in interactive sessions as well as a batch task.

MATLAB (matrix laboratory) is a high-level programming language and interactive environment for numerical computing, visualization, and programming that was developed in the fourth generation.

It supports matrix manipulations, function and data charting, algorithm implementation, user interface design, and interfacing with programme written in other languages such as C, C++, Java, and FORTRAN. It also helps you to analyze data, build algorithms, and create models and applications.[33]

The following are some of the most regularly used mathematical calculations:
- Working with Matrices and Arrays
- 2-D and 3-D Plotting and Graphics
- Linear Algebra
- Algebraic Equations
- Non-linear Functions
- Statistics
- Data Analysis
- Calculus and Differential Equations
- Transforms
- Curve Fitting
- Various additional specialized operations

B. R Software

R is a free statistical computing and graphics software environment. It builds and operates on a broad range of UNIX, Windows, and MacOS systems.

In the fields of quantitative analysis and machine learning, the R programming language is a critical chevalier. The language's popularity is only projected to rise as machines become increasingly significant as data producers. R, on the other hand, offers both advantages and disadvantages that developers should be aware of.

R was initially introduced in the 1990s and has functioned as an implementation of the S statistical programming language, as seen by language popularity indices such as TIobe, PyPL, and Redmonk. The easiest method to generate repeatable, high-quality analysis is to use R.

R's benefits in machine learning are mostly due to R's strong links to academia.

The fundamental idea behind R comes from 1960s programming languages.

R also can't be loaded into a browser. Because of its lack of security via the Web, using R as a back-end server for computations was almost impossible. However, advances such as the utilization of virtual containers on the Amazon Web Services cloud platform have helped to mitigate the security issue.[34]

C. TORA: The Temporally Ordered Routing Technique (TORA) is a data routing algorithm that may be used in Wireless Mesh Networks or Mobile Ad Hoc Networks. Vincent Park and Scott Corson of the University of Maryland and the Naval Research Laboratory created it. Park's technique has been patented, and Nova Engineering has licensed it to commercialize a wireless router based on Park's algorithm.

The TORA uses a "flat," non-hierarchical routing method to try to achieve great scalability. The algorithm's operation aims to suppress the formation of far-reaching control message propagation to the maximum extent feasible. The TORA does this by avoiding the use of a shortest path solution, which is rare for routing algorithms of this sort.

TORA creates and maintains a destination-rooted Directed Acyclic Graph (DAG). The height of nodes cannot be the same.

The localization of control messages to a relatively limited selection of nodes near the occurrence of a topological change is one of TORA's fundamental design elements. Nodes must preserve routing information about neighboring (one hop) nodes in order to do this. The protocol has three primary functions:
- Creating a route
- Routing and maintenance
- Erasure of routes

Because the height measure is reliant on the logical time of the connection breakdown, timing is critical for TORA.[35]

D. Hadoop: Hadoop is an open-source software framework for storing and processing data on commodity hardware clusters. It has a lot of storage for any type of data, a lot of processing power, and the capacity to perform almost unlimited concurrent processes or jobs.[36]

Significance of Hadoop
- The ability to swiftly store and handle large volumes of any type of data. That's an important concern as data quantities and kinds continue to grow,

notably from social media and the Internet of Things (IoT).

• Processing power -Hadoop's distributed computing paradigm efficiently handles large amounts of data. You have higher processing power if you use more computer nodes.

• Tolerance to faults - Hardware failure does not affect data or application processing. Jobs are automatically routed to other nodes if a node fails, ensuring that distributed computing does not fail. All data is automatically duplicated and saved in many locations.

• Flexibility - We don't have to preprocess data before saving it, unlike standard relational databases. You may save as much information as you like and decide how to utilize it afterwards. Text, photos, and videos are examples of unstructured data.

• Low cost - The open-source framework is free and stores massive amounts of data on cheap hardware.

• Scalability - By simply adding nodes, you may quickly expand our system to accommodate additional data. There is very little administration required.

F. Nosql Database: It focuses on storing and retrieving massive amounts of organized, semi structured, and unstructured data. It ensures read-write reliability. NoSQL databases are more scalable and give higher performance than relational databases, and its data format tackles certain difficulties that the relational model is not meant to address: large amounts of organized, semi-structured, and unstructured data that is constantly changing.[37]

G. Map Reduce: MapReduce is a distributed computer processing technology and programme architecture based on Java. Map and Reduce are two fundamental jobs in the MapReduce algorithm. Map turns a collection of data into another set of data by breaking down individual pieces into tuples (key/value pairs). Second, there's the reduction job, which takes the result of a map as an input and merges the data tuples into a smaller set. The reduction work is always executed after the map job, as the name MapReduce suggests.

MapReduce's main benefit is that it's simple to expand data processing over several computer nodes. Mappers and reducers are the data processing primitives in the MapReduce paradigm. It might be difficult to break down a data processing application into mappers and reducers. Scaling an application to operate over hundreds, thousands, or even tens of thousands of servers in a cluster is only a configuration modification once we build it in MapReduce style. Many programmers have been drawn to the MapReduce approach because of its straightforward scalability.[38]

H. Hive: Hive is a Hadoop data warehouse architecture solution that allows you to handle structured data. It sits atop Hadoop to summarize Big Data and facilitate searching and analysis. Initially created by Facebook, Hive was eventually taken up by the Apache Software Foundation and maintained as an open-source project under the name Apache Hive. Various businesses make use of it. Amazon utilizes it in Amazon Elastic MapReduce, for example.

Hive is used to process structured data. It's a data warehouse design. Hive is mostly used for writing SQL Scripts and is not a relational database.[39]

I. Pig: Pig is a high-level programming language for Apache Hadoop. Pig allows data analysts to create complicated data transformations without having to know Java. Pig Latin is a basic SQL-like programming language that appeals to developers who are already comfortable with scripting languages like SQL. Pig is complete, so we can use it to perform all of the essential data transformations in Apache Hadoop. Pig processes data from a variety of sources, both structured and unstructured, and stores the findings in the Hadoop Data File System. Pig scripts are converted into MapReduce tasks, which are then executed on the Apache Hadoop cluster.

Pig is the outcome of Yahoo's development efforts. Pig allows users to spend more time studying large data sets and less time building Map-Reduce scripts. Pig programming language is meant to operate with any type of data, similar to how pigs devour anything. That's how Pig got his name! It is a "Perl-like" language, rather than a "SQL-like" language, that enables for query execution on data throughout the cluster. It is an open-source framework as well.[40]

J. Skytree: SkyTree is a storage system. It is a high-performance machine learning and data analytics platform. Rather of manually gathering data, machine learning is more effective.

High-tech tools are currently upping the standard, allowing farmers to analyze vast quantities of data acquired through sensors to anticipate the optimum time to plant, what sort of seed to use, and where to plant in order to increase yields, save operating costs, and reduce environmental impact. Farmers may use systems like John Deere's Farm Sight, Monsanto's Field Scripts, and Pioneer's Field360 to collect planting and yield data from monitored farm equipment and enter it into a database that, when combined with several sources of anonymized data, generates precise prescriptions.[41]

K. Spark and Spark Stack: The RAD Lab at UC Berkeley created Spark (now called as AMP Lab). Matei Zaharia et al. [42],[43] are the key contributors. Its original goal was to turn Hadoop into a general-purpose framework that used resilient distributed datasets (RDDs) and micro batch processing in memory. To put it another way, it aims to replace a MapReduce model with a better one. It focuses on the computing efficiency of iterative and recursive algorithms, as well as interactive data mining queries. It claimed that for specific workloads, such as conducting iterative algorithms, it would be 10–20 times quicker than MapReduce. It does not abandon HDFS in favor of Hadoop's file storage system, despite its attempts to replace MapReduce.

It is an open-source project under the Apache Software Foundation, like many other Hadoop-related projects. It was relocated to ASF as an incubator in June 2013. It has been an Apache top-level project since 2014, and numerous Big Data providers, including Cloudera, Horton, SAP, and MapR, support it.

Spark is a general-purpose processing platform built on huge clusters that is fast and scalable. In contrast to MapReduce, which is primarily developed for a web crawler, indexing system, and restricted machine learning, Spark's compute platform incorporates SQL, interactive query, data stream, graph, and machine learning analytic features. Spark was built as a single stack based on the Berkeley data analytics stack architecture, unifying all libraries and higher-level components.

Spark is made up of seven main components: a data engine, a cluster manager (which includes Hadoop, Apache Mesos, and a built-in standalone cluster manager), Spark SQL, Spark streaming, Spark Machine Learning Library, Spark GraphX, and Spark programming tools.

IX. PROBLEMS FACED BY THE AGRICULTURE SECTOR & PRESENT AGRICULTURE SYSTEM IN INDIA [44]

A. *Small and fragmented land-holdings*

Varying tracts have different fertility values and should be dispersed correspondingly. If there are four tracts to be divided between two sons, each son will receive a lesser portion of each tract. With each passing generation, the holdings get smaller and more dispersed. One of the key causes of our low agricultural output and backward position is the subdivision and fragmentation of holdings. Moving seeds, manure, equipment, and animals from one piece of land to another wastes a lot of time and effort. Small land holdings limit the farmer's ability to adopt conventional farming practices and reduce production. Because land holdings are tiny, more people are forced to labour on farms in rural regions, and agricultural revenues suffer as a result of outdated technologies.

One of the main reasons why agriculture isn't job-friendly is that it only makes farming lucrative for a small percentage of farmers.

B. *Seeds*

Seed is an essential and basic component for increasing crop yields and maintaining agricultural output growth. The distribution of high-quality seed is just as important as its manufacturing. Unfortunately, excellent quality seeds are out of reach for the majority of farmers, particularly small and marginal farmers, due to excessive seed prices.

The Government of India formed the National Seeds Corporation (NSC) in 1963 and the State Farmers Corporation of India (SFCI) in 1969 to address this issue. Thirteen State Seed Corporations (SSCs) were also created to help farmers get access to better seeds.

Breeder, foundation, and certified seeds are the three types of generation recognized by the system. The basic seed and initial step in seed manufacturing is breeder seed. The progeny of breeder seed, foundation seed is the second step in the seed production cycle.

C. *Manures, Fertilizers and Biocides:*

Over thousands of years, Indian soils have been utilized to cultivate crops with little regard for replenishment. As a result, soils have been depleted and exhausted, resulting in low production.

Manures and fertilizers serve the same purpose for soils as adequate nutrition does for the human body. A well-nourished soil may provide good crops in the same way as a well-nourished body can accomplish any decent work.

As a result, a rise in fertilizer usage is a sign of agricultural prosperity. Cow dung is the most effective manure for soils. However, because so much cow dung is utilized as kitchen fuel in the form of dung cakes, its usage as such is restricted. Chemical fertilizers are expensive and sometimes out of reach for small farms. As a result, the fertilizer issue is both acute and complex.

Organic manures are thought to be necessary for maintaining the soil's health. The country has a compost potential of 650 million tones for rural use and 160 lakh tones for urban use, which is currently underutilized. The harnessing of this potential will tackle the dual problems of garbage disposal and soil fertilization.

Pests, viruses, and weeds cause significant crop loss, accounting for almost one-third of total field yield at the time of Independence. To conserve crops and reduce losses, biocides (pesticides, herbicides, and weedicides) are utilized.

D. *Irrigation*

Even though India is the world's second-largest irrigated country after China, irrigation covers just one-third of the planted land. In a tropical monsoon nation like India, where rainfall is unpredictable, inconsistent, and erratic, irrigation is the most critical agricultural input. India will not be able to make sustained development in agriculture unless and until more than half of the planted area is irrigated.

However, extreme caution must be exercised to avoid the negative impacts of excessive irrigation, particularly in regions watered by canals. Due to improper irrigation, large expanses of Punjab and Haryana have been rendered worthless (areas afflicted by salinity, alkalinity, and water logging). The majority of farming in India is monsoon-dependent; when the monsoons are good, the whole economy (not just the agricultural sector) thrives; when the monsoon fails, everyone suffers. The issue here is either a shortage of or improper management of water.

E. Sustainability problems

Due to the usage of outmoded farming technologies, Indian agricultural production is considerably below international norms. In addition, the impoverished agricultural community's lack of knowledge of the need for sustainability has exacerbated the situation. Some desert areas abuse the irrigation infrastructure given by growing water-intensive crops, resulting in unexpected water use. Ground water resources are frequently used in locations where irrigation in the form of rivers and canals is insufficient. Agriculture's sustainability is critical since many of the difficulties farmers encounter are linked to it.

Farmers will be pushed into a vicious spiral of loans, high fertilizer usage, water mismanagement, low productivity, and thus more debt as a result of a lack of knowledge of the need to cultivate crops sustainably.

F. Lack of mechanization

Despite the fact that agriculture has been heavily mechanized in some regions of the nation, the majority of agricultural activities are still carried out by hand utilizing simple and traditional equipment and implements such as the wooden plough and sickle.

Ploughing, seeding, irrigating, thinning and pruning, weeding, harvesting threshing, and transporting the crops are all done by hand. This is particularly true for small and marginal farmers. It leads to significant waste of human labour and poor labour yields per capita.

It is critical to mechanize agricultural activities in order to reduce labour waste and make farming more convenient and efficient. Agricultural tools and machinery are essential inputs for effective and timely agricultural operations, allowing for multiple cropping and hence increased output.

The increased use of tractors, power tillers, combine harvesters, irrigation pumps, and other power powered devices contributed to this growth. Mechanical and electrical power now account for 84 percent of total energy, up from 40% in 1971.

G. Soil erosion

Large areas of productive land are subjected to wind and water erosion. This region must be handled appropriately and returned to its former fertility.

H. Agricultural Marketing

Storage facilities in the rural areas are either totally absent or grossly inadequate. Under such conditions the farmers are compelled to sell their produce immediately after the harvest at the prevailing market prices which are bound to be low. Such distress sale deprives the farmers of their legitimate income.

At present there are number of agencies engaged in warehousing and storage activities. The Food Corporation of India (F.C.I.), the Central Warehousing Corporation (C.W.C.) and State Warehousing Corporation are among the principal agencies engaged in this task. These agencies help in building up buffer stock, which can be used in the hour of need.

This scheme provides storage facilities to the farmers near their fields and in particular to the small and marginal farmers. The Working Group on additional storage facilities in rural areas has recommended a scheme of establishing a network of Rural Storage Canters to serve the economic interests of the farming community.

The government has created regulated marketplaces in order to protect farmers from the clutches of money lenders and middlemen.

I. Inadequate storage facilities

In remote places, storage facilities are either non-existent or woefully inadequate. Farmers are forced to sell their crops soon after harvest at market prices, which are almost always poor. Farmers lose their rightful revenue as a result of such distress sales.

There are now a lot of companies involved in warehousing and storage. The Food Corporation of India (F.C.I.), the Central Warehousing Corporation (C.W.C.), and the State Warehousing Corporation are some of the major players in this endeavor. These organizations assist in the development of a buffer stock that can be used in an emergency.

Farmers near their fields, particularly small and marginal farmers, benefit from this initiative, which offers storage facilities. A concept to construct a network of Rural Storage Canters to serve the economic interests of the agricultural community has been suggested by the Working Group on additional storage facilities in rural regions.

J. Inadequate transport and Supply channel bottlenecks and lack of market understanding

One of the most significant challenges in Indian agriculture is the absence of affordable and effective transportation. Thousands of communities are still without access to big roadways or market centers.

The majority of rural roads are Kutcha (bullock-cart roads), which become impassable during the wet season. Farmers are unable to get their goods to the major market and are compelled to sell it at a cheap price in the local market.

Farmers are forced to distress sales because to a lack of a competent marketing route, making them prey of greedy intermediaries and limiting their revenue.

India produces about 265 million tons of food grains each year, more than enough to feed its entire population for several years. Despite this, we observe a lot of wasted food, rising food prices, and millions of hungry people. This must come to an end.

K. Scarcity of capital

Agriculture is a significant business that, like all others, needs money. With the progress of farm technology, the function of capital input is becoming increasingly crucial. Because the agriculturist's capital is tied up in his fields and stocks, he must borrow money to keep agricultural output moving.

Moneylenders, dealers, and commission agents are the primary sources of funds for farmers, charging high interest rates and purchasing agricultural goods at a low cost.

The rural finance landscape has changed dramatically, with institutions such as Central Cooperative Banks, State Cooperative Banks, Commercial Banks, Cooperative Credit Organizations, and certain government agencies providing farmers with low-interest loans.

Agriculture is one area in which India excels above others, despite the fact that it has been deemed a dead sector for several decades.

L. Overdependence on traditional crops like rice and wheat

To get the highest yields, each crop requires certain climatic conditions. Despite the fact that rice and wheat are grown in broad parts of India, particular places may easily convert to other crops to increase output. Even though we have the requisite circumstances to cultivate more oilseeds here, India imports cooking oil from elsewhere.

The lack of a robust national agriculture strategy is seen by the heavy reliance on traditional rice and wheat. Excess inventories in a few crops cause issues with selling, storage, and the availability of other farm products.

Furthermore, if farm output is biased toward crops like rice, farmers would overuse irrigation and ground water resources, resulting in a slew of additional issues.

X. AGRICULTURE AND BIG DATA

Agriculture is by far one of India's strongest assets. India is noted for having a favorable ratio of arable land to other countries. India, being the world's second most populous country, has the highest ratio of land to population density. Rapid urbanization has squid the ratio of cultivable/agricultural land ownership. Increased investments in agricultural infrastructure such as irrigation facilities, warehousing, and cold storage are predicted to boost the Indian agriculture sector in the coming years.

In addition, the increased usage of genetically modified crops is expected to boost yields for Indian farmers. A reliable and controllable IT infrastructure is also necessary to enable all of this and more. India has lagged behind Israel, Canada, and the United States in terms of using IT and ICT in agriculture.

Big Data has been utilized in many nations throughout the world to unlock untapped potential in order to increase the productivity, profitability, and competitiveness of farmers and ranchers, ultimately preserving natural resources and improving the environment. In reality, the advantages of Big Data have already begun to pay off in all nations where IT and ICT use has been adapted to accept Big Data and its Analytics.

A. Big Data matters

The vast volume of data – both organized and unstructured – that inundates a firm on a daily basis is referred to as big data. But it's not the quantity of data that matters. What matters is what companies do with the data. Big data may be studied for insights that lead to improved business choices and strategic movements.

B. Importance of Big Data in Agriculture of India

The value of big data is determined by what we do with it, not by how much data we have.

We can achieve agriculture-related tasks by combining big data with high-powered analytics, such as:

• The soil fertility database, which decides which crop produces the best results, is first and foremost.

Soil conditions for legumes, millets, and oil seeds, for example, vary. Soil fertility databases will undoubtedly be the starting point at the state and national levels.

• The second stage is to collect consumption metrics for each food grain; this data will aid in determining closeness so that each food grain may reach its ultimate user with the least amount of transportation.

• The third phase is to rely on farm-saved seed rather than corporate seed monopolies, which will result in a paradigm change in farming that leads to organic farming.

The fourth stage is to register the traders and determine the metrics for kind, quantity, and price.

• The fifth stage is to map the end user and determine who he is and how much he consumes each month or year. We can automate the data acquired by retail shops, which are already issuing invoices for each item purchased.

• The mapping of the climate prediction system is the sixth phase.

• Mapping local farmers and traders with the export community.

• In near-real time, determining the fundamental causes of failures, difficulties, and faults.

• Complete risk portfolios may be recalculated in minutes.

C. The journey ahead

The most critical aspect of moving PA ahead will be the development of a large pool of engineers, scientists, and agriculturists to work on various aspects of the technology. PA will not flourish without strong staff and, as a result, good R&D. Unfortunately, the majority of talented students desire to pursue careers in engineering or medicine. Agriculture is relegated to the background. Excellent engineers from IITs, NITs, and other schools are needed to create machinery for PA, such as robots and drones. This can be aided by the creation of a new area of engineering called agricultural mechanotrics or robotics, in which academics and students from nearly all branches of engineering will interact and collaborate to design smart systems for PA.

"Excellent engineers from IITs, NITs, and other schools are needed to create machinery for PA, such as robots and drones. Establishing a new field of engineering termed agricultural mechanotrics or robotics will help with this..."

Another option is to collaborate with scientists from ICAR institutions, engineers, industry, and farmers to create PA. I believe that industries must take the lead since they will create the machinery and establish the leasing agency. With more employment being created in Pennsylvania, more students will want to pursue a career in agriculture.

The information gathered can be utilized in the following scenarios.

1. Weather trends in the past
2. Information on plant breeding and production for each strain
3. Pesticide and fertilizer standards
4. Information on soil production
5. Information about water supply
6. Data on market spot prices and futures

Fig.1- Capturing data

Fig.2 –Live analyst of data

Fig. 3- Sensor for climatic changes detection

D. Sensors used in data collecting:

The various sensors used in data gathering are given below.

The following are some of the advantages that sensors and drones bring.

• Field and resource mapping using sensors

•Crop monitoring from afar Remote crop monitoring

•Climate monitoring and forecasting

•Weather conditions may be forecast with ease.

•It is possible to reduce the use of fertilizers and pesticides.

Sensors can determine how successful different seeds and fertilizer kinds are in different parts of a farm. For maximum production, the software will direct the farmer to plant one hybrid in one corner and a different seed in the other. It may modify nitrogen and potassium levels in various sections of the soil.

Some of the applications below can be seen in a particular manner of collecting data.

Fig.4- Sensor for pesticide control detection

Fig. 5 – Sensor for diseases detection

Fig 6 – Sensor for Moisture Control

Fig: 7 - Using drones to capture image

Measuring agricultural land quality to determine field productivity (Vehicle Mounted Sensors)

Creating vehicle-mounted sensor systems for spatial and temporal analysis of the land is a nice notion that may be implemented for India to benefit farmers. This will assist to reduce the time it takes to conduct soil analysis, which is presently done by obtaining soil samples and shipping them to a lab. Furthermore, vehicle-mounted sensors can evaluate more samples, resulting in a more accurate soil analysis.

Government organizations may construct such vehicles and rent them to farmers to help them increase output in exchange for real-time data on predicted productivity (predictive) from various places.

Many sensors can be implemented on vehicle mounted interfaces

A. *Electrochemical sensors:* The most significant sensors are those that measure soil nutrient levels and pH.

B. While they may not be as exact as laboratories, they may be utilized to produce relatively accurate readings in real time for large samples, which is important for decision-making and agricultural automation.

C. *Electromagnetic sensors:* for Soil pH, Salinity, Moisture etc,

D. *Optical sensors:* To determine organic matter, soil type, and moisture content. Several. When gazing at soil, these sensors may replicate the human eye and assess near-infrared, mid-infrared, or polarized reflectance.

E. *Mechanical sensors:* Such sensors are currently available on tractors to regulate traction control due to soil compaction.

F. *Airflow sensors:* for measuring permeability.

Drones can give three different sorts of detailed images to farmers. For starters, looking at a crop from above can reveal patterns that suggest everything from irrigation issues to soil variance to insect and fungal infestations that aren't visible at eye level.

Second, aerial cameras may capture multispectral pictures, which combine data from the infrared and visible spectrums to generate a crop view that shows distinctions between healthy and distressed plants in a way that can't be seen with the human eye.

Finally, a drone may inspect a crop weekly, daily, or even hourly. When those images are combined to make a time-series animation, they can display changes in the crop, exposing difficulty locations or chances for crop improvement. Agriculture drones are becoming towards tools similar to any other consumer item, and we are beginning to discuss what we can do with them.

XI. IMPLEMENTATION OF BIG DATA ANALYTICS IN AGRICULTURAL SYSTEMS

We have now entered the Big Data era. Big data provides a platform for acquiring, storing, and analyzing data in order to uncover previously unknown knowledge.

Sensor-based solutions are being investigated and deployed in various areas of agriculture across the world.

By judiciously utilizing ever-maps and data, moisture content maps and data, air permeability maps and data, nutrient contents and pH level data, past cultivation records, insurance and yield-related information, and social media posts such as tweets, blogs, new feeds, and articles in agriculture journals By properly assessing the numerous aspects, the big data scientist's role is to mine the big data and uncover relationships, patterns, and trends in order to enhance agricultural systems, boost crop yield, and cut costs.

The majority of agricultural data analytics are performed using yearly data recreation in relational databases that provide pre-processed reports. The data should be analyzed immediately.

Furthermore, data vivification must be performed in real time rather than once a month or once a year.

A variety of sensors can be utilized to manage variable rate application equipment in real-time or to produce field maps of specific soil attributes using a Global Positioning System (GPS).

The agricultural system must always adapt and innovate in order to provide greater services to the people. The information obtained from big data analysis may be used in a variety of ways for precision agriculture.

Some of them are as follows:

- The data collected by multiple systems (or sensors) may be analyzed in real-time to indicate essential values that are crucial in manufacturing decision-making.
- The resulting geographical maps have a very high resolution and may spot variations in soil moisture. It would point to the strategic use of irrigation. With the use of detailed photographs of insect damage in the field, precise control targeting will also be achievable.

These big data apps can be quickly tested, refined, and improved, and they will completely transform agricultural delivery and research. Though big data analytics in agriculture is important for providing improved agricultural services, it also enables historical data analysis to unearth undiscovered facts. Data heterogeneity and incompleteness, scalability, timeliness, privacy, and human cooperation are key issues in big data analytics. The goal of future research is to overcome hurdles and apply big data analytics in agriculture to uncover proficiency from unstructured data.

XII. CONCLUSION AND FUTURE WORK

The most major criticism of mechanized agriculture is that agricultural machinery is prohibitively expensive for most farmers. However, establishing agricultural equipment leasing firms in rural regions might be a viable solution. These businesses will rent automated equipment to farmers, including drip irrigation systems, as well as providing skilled labour to operate these devices. On a small basis, this already happens in India, where certain agencies harvest wheat using combines and spray crops. Farmers are charged on an hourly basis.

With a scarcity of agricultural labour, it is an appealing and cost-effective proposition for many farmers.

Critics of automation also argue that a farmer may simply boost agricultural output by planting crops on time and providing adequate and suggested water and fertilizer. However, the availability of labour, water, and fertilizer determines how well and when inputs are applied. This isn't always practicable, especially when many farmers rely on the weather gods for irrigation and labour is in short supply. Precision agriculture can assist with this.

A farmer's ultimate goal should be to find superior crops, use the seed to propagate them further, and therefore become a breeder.

Progressive farmers are already doing so, and with additional time due to PA, they may be able to create better and higher producing cultivars. Mechanization will also make crop producing appear less scary, thereby attracting a larger number of individuals to take up farming.

As a result, high-tech PA can assist in delivering India's next green revolution and producing significant rural prosperity in a sustainable and ecologically sound manner.

REFERENCES

[1]. https://en.wikipedia.org/wiki/*Agriculture_in_India*
[2]. Rigveda - Wikipedia
[3]. Agriculture's share in GDP declines to 13.7% in 2012-13
[4]. "CIA Factbook: India". CIA Factbook. Central Intelligence Agency. Archived *from the original on 11 June 2008* Retrieved 2008-06-10.
[5]. https://pib.gov.in/PressReleasePage.aspx?PRID=1814057 & https://www.pib.gov.in/PressReleasePage.aspx?PRID=1741942
[6]. India's Agricultural Exports Climb to Record High, United States Department of Agriculture (2014)
[7]. Tekiner, F.; Keane, J.A., "Big Data Framework," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, vol., no., pp.1494, 1499, 13-16 Oct. 2013
[8]. http://www.acquia.com/examples-big-data-projects.
[9]. https://statisticstimes.com/economy/country/india-gdp-sectorwise.php
[10]. https://en.wikipedia.org/wiki/Economy_of_India#cite_note-153
[11]. Smith TP. How big is big and how small is small, the size of everything and why. USA: Oxford University Press; 2013 p. 14–29.
[12]. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking- index-vni/ VNI_Hyperconnectivity_WP.html.
[13]. Gil Press, "A Very Short History of Big Data," Forbes Tech Magazine, May 9, 2013. URL:http://www. forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/.
[14]. Rider F. The scholar and the future of the research library. A problem and its solution. New York: Hadham Press; 1944.
[15]. Ohlhorst F. Big data analytics, turning big data into big money. Canada: John Wiley & Sons, Inc; 2013 p. 2.
[16]. http://www.winshuttle.com/big-data-timeline/
[17]. http://visualizing.org/visualizations/big-data-brief-history.
[18]. Laney D. 3D data management: controlling data volume, velocity and variety. USA:Application Delivery Strategies, Meta Group; 2001. p. 1–4.
[19]. http://www-01.ibm.com/software/data/bigdata/.
[20]. http://www.ibmbigdatahub.com/infographic/four-vs-big-data.
[21]. Zikopoulos PC, et al. Harness the power of big data, the IBM Big data platform. US: McGraw-Hill; 2013. p. 9
[22]. www.microsoft.com/bigdata.
[23]. Demchenko Y. Defining architecture components of the big data ecosystem. In: IEEE Collaboration Technologies and System (CTS); 2014. p. 104–12.14501649, ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6867550&url= http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6867550, US.
[24]. http://timoelliott.com/blog/2013/07/7-definitions-of-big-data-you-should-know-about.html.
[25]. Pearce R. "Big data is BS: Obama campaign CTO," CIO Magazine, May 28, 2013.http://www.cio.com.au/ article/462961/big_data_bs_obama_campaign_cto/
[26]. http://www.ingrammicroadvisor.com/data-center/four-powerful-big-data-application- examples
[27]. G. Rajesh Kumar **et al., (2015):** 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy, and Materials, "Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique" (ICSTM).
[28]. Rupika Yadav et al. 2015. Big Data Meets Small Sensors in Precision Agriculture, International Journal of Computer Applications (0975 – 8887) Applications of Computers and Electronics for the Welfare of Rural Masses (ACEWRM) 2015.
[29]. Hemlata Channe, et al., XXXX. Multidisciplinary Model for Smart Agriculture using Internet-of-Things (IoT), Sensors, Cloud-Computing, Mobile-Computing &Big-Data Analysis, Int.J.Computer Technology &Applications, 6(3): 374-382.
[30]. Duncan Waga et al., 2014. Environmental Conditions' Big Data Management and Cloud Computing Analytics for Sustainable Agriculture, World Journal of Computer Application and Technology, 2(3): 73-81.
[31]. Ling Zang et al. 2009. Research on the AID Measurement Based on Improved SVM Algorithm, Second International Symposium on Electronic Commerce and Security.
[32]. https://en.wikipedia.org/wiki/Weka_(machine_learning)
[33]. www.tutorialspoint.com/matlab
[34]. https://www.r-project.org/about.html

[35].https://en.wikipedia.org/wiki/Temporally_ordered_routing_algorithm
[36]. http://www.sas.com/en_us/insights/big-data/hadoop.html
[37]. https://www.mongodb.com/nosql-explained
[38]. http://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm
[39]. http://www.tutorialspoint.com/hive/hive_introduction.htm
[40]. http://www.guru99.com/introduction-to-pig-and-hive.html
[41]. https://datafloq.com/read/john-deere-revolutionizing-farming-big-data/511
[42] Zaharia M, et al. Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing. In: Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, April 25–27; 2012.
[43] Karau H, Konwinski A, Wendell P, Zaharia M. Learning spark. USA: O'Reilly Media Inc; 2015
[44]. http://www.yourarticlelibrary.com/agriculture/10-major-agricultural-problems-of-india-and-their-possible-solutions/20988/