A case study on :

# DATA SCIENCE IN AGRICULTURE


**Prepared by** : Brijesh Rameshbhai Rohit

**Roll No.** : U19CS009

**Class** : B.Tech-III Computer Science and Engineering

5th Semester

**Year** : 2021-22

**Subject** : Data Science

**Department of Computer Science and Engineering**

**Sardar Vallabhbhai National Institute of Technology,**

**Surat -395007 (Gujarat), India**

# Table of Contents

# List of Figures

# Chapter 1

# Abstract

The effect of big data is immense in every business area. The utilization till now has been targeted at the e-commerce and marketing sectors. However, the massive reach of big data can provide much more innovative, worthwhile, and valuable answers for plenty of perennial troubles faced by various sectors. Rural India performs an essential role in the nation's economic boom through agriculture, self-employment, construction, and so forth.

The advancement in the telecommunication sector provides an appropriate platform for implementing data analytics in rural India. The focus of the government to build digital India through broadband highways connecting each family, village, panchayat, etc., will generate a large number of records that can be analyzed to offer a solution to the never-ending problems of rural India and create smarter villages. Sustainable agricultural development is a significant solution with fast population development through the use of information and communication (ICT) in precision agriculture, which produced new methods for making cultivation further productive, proficient, well-regulated while preserving the climate. Big data (machine learning, deep learning, etc.) is amongst the vital technologies of ICT employed in precision agriculture for their huge data analytical capabilities to abstract significant information and to assist agricultural practitioners to comprehend well farming practices and take precise decisions.

This paper presents a conceptual framework for the application of data analytics in enhancing rural development by using assisting unique sectors consisting of agriculture, banking, governance, and healthcare. This article features data creation methods, accessibility of technology, accessibility of devices, software tools, and data analytic methods, and appropriate applications of big data in precision agriculture. Besides, there are still a few challenges that come across the widespread implementation of big data technology in agriculture.
***Keywords—*** Precision agriculture, big data analytics, machine learning, sustainable agriculture, smart farming, and digital agriculture.

# Chapter 2

# Introduction

With the development in computer technology and the growing digitization of services, citizens benefit from quality services worldwide. However, massive electronic data are being generated within the procedure of digitization of public services. Smartphones, video recordings, smartphone applications like WhatsApp, FB, Twitter, and so on. Connecting to the internet produces massive records. In this context, the advancement of Big Data pertains to modern techniques of accumulating, storing, and processing data to grow its availability and usability. Big Data is the term that describes the massive volume of facts, both structured and unstructured, that affect an organization on an everyday basis. Big data may be analyzed for insights that result in strategic decisions making inside the businesses.

Gathering and storing information is through an age-old concept, the concept of Big data is relatively new, gained popularity only in early 2000s when industry analyst Doug Laney framed the present definition of Big data as three Vs:

1. Volume: Data are collected from a variety of sources-business transactions, social media, mobile data and information from sensors and so on. Storing data was a challenge in the past but new technologies have reduced this burden

2. Velocity. Data keep on multiplying with extreme speed and hence must be dealt in timely manner. Sensors, login machines in the offices all drive in the need to deal with large volume of data in limited time

3. Variety: Data are found in different types and formats from structured, numeric data to unstructured text documents like email, audio, smart messages. The other two dimensions identified are variability and complexity as the data accumulated are found to be more diverse and complex in nature.

To deliver sustainable agriculture production, the agriculture sector needs to employ cutting-edge technologies like blockchain, IoT, and AI. Data-driven agriculture with these technologies is the most promising approach to solve existing and future problems. If we could generate a huge quantity of data from the farm and use that data to drive some of the agricultural decisions. It can help to solve most of these food problems globally. For instance, if we could enable farms to build data sets or maps for soil moisture, temperature and humidity in the area, availability of water, and other environmental factors around the farm, itwould enable techniques like smart farming, precision agriculture, vertical farming, etc.

Data-driven agriculture has been shown to improve crop yield, reduce cost, and ensure sustainability [1]. These are not limited to agriculture but have potential solutions for several challenges faced by livestock farming also.

Digitanimal is a company to enhance livestock farm productivity, sustainability, and animal welfare along with providing thorough monitoring solutions based on IoT wearable's powered by firmware, AI, satellite images, and blockchain technology provides farmers with relevant information on health, location, feeding, and reproduction conditions of their animals. Thus by incorporating technological-based agriculture techniques increases their yield, reduces costs, and improves the income of farmers, thereby improves their quality of life. Smart agriculture deems it essential to address these problems that have attracted a lot of technological attention, from sowing to watering of crops to health, harvesting, and traceability in supply chain management.

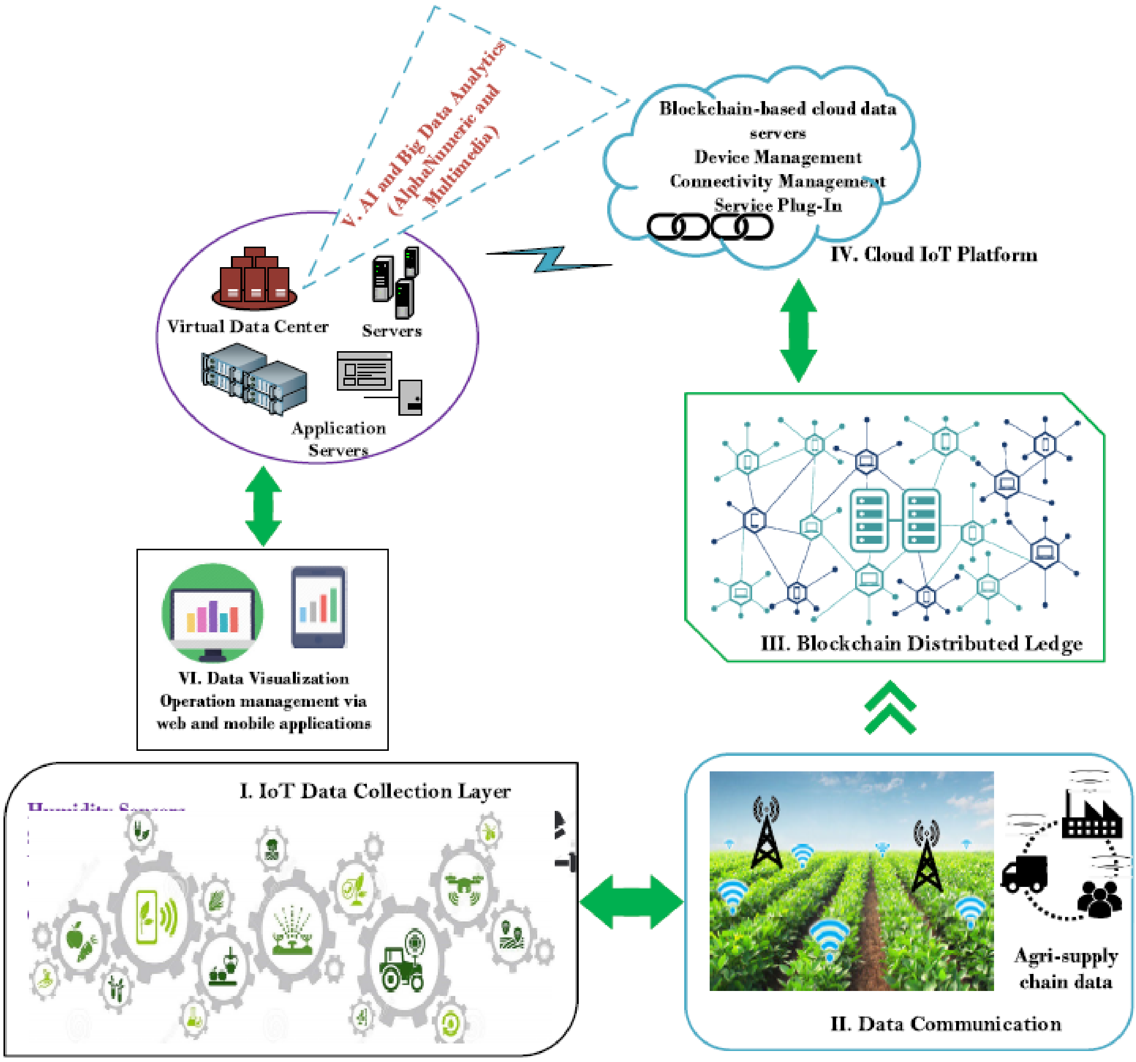


**Figure 2.0.1:** Smart Agriculture Ecosystem

Figure 2.0.1 shows an overview of these technologies concerning the smart agriculture ecosystem. Big data empowers agricultural practitioners and related industries to gain information about different factors that influence agricultural production and take efficient decisions in daily farming. It keeps them up to date about the market price, demand of a particular crop, and the new technologies in the agriculture sector. Recently, big factory farms have embraced different technologies like IoT and blockchain with an intent to produce greater production in the farming practice. Blockchain technology is being implemented in the management of the agri-food supply chain to make available features such as transparency, security, immutability, and reliability of all operations. Blockchain also assists in addressing several IoT security and reliability challenges. IoT assists in data collection at each stage of agriculture production and supply chain [2]. Therefore, it would likewise be valuable to perform big data analytics on the data collected during farming, processing, logistics, and marketing.

# Chapter 3

# Big Data

Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/ or analysis. From a technology perspective, big data gives: the possibility of better storage (volume), the ability to process the information and make it available in real time (velocity) and the ability to deal with various kinds of data sources, including structured, semi-structured and unstructured ones (variety).

Data could be coming from disparate sources like Geographical Positioning System (GPS) devices, sensors, social networking sites, company websites, traditional, legacy data sources etc. Assembling such wide and deep data at single place, integrating it and gaining meaningful understanding is a mammoth task. The impact of big data is enormous virtually in every business sector. The usage so far has been more focused on the e-commerce and marketing sectors. But the wide reach of big data can provide much more innovative, profitable and yet beneficial solutions for many perennial problems faced by different segments.

Big data in agriculture has the potential to increase yield production, and as we near an era shaped around more people and less resources, this makes farming one of the most important careers in the world. Many public and private companies are increasingly relying on BI and decision support tools and services to obtain better metrics from their current systems, make more timely and informed decisions, and bring about more transparency and accountability. Big data offers opportunity for improved understanding of human behaviour that can facilitate early warning, real-time awareness and feedback. Finding an implementation solution for big data is often difficult as it should combine scalability, performance, ease of use and low total cost of ownership. Challenges in the application of big data are mainly related to privacy, access and sharing of data.

## 3.1 Big Data Analytics

Big data analysis is outlined as a system in which cuttingedge analytic methods operate on huge data sets. Therefore, it is a combination of two technical entities massive amount of data sets, and a collection of analytics tool categories including data mining, statistics, AI, predictive analytics, natural language processing (NLP), etc. forming an important component of business intelligence.


**Figure 3.1.1:** Big Data Characteristics.

Big data is depicted by the subsequent attributes which are shown in figure 3.1.1. Big data is being used in numerous fields such as big services business industries like Amazon to learn customer behavior and needs more precisely to tailor product prices accordingly, enhance operational productivity, and cut down personal costs. Even social networking sites Facebook, Twitter, and other networking sites utilize big data analytics to study your social behavior, interests, and social connections and then endorse the specific products. In an intelligent transportation system, big data techniques can handle the enormous quantity of diverse and complex data generated over the period to provide safe and superior facilities aimed at drivers and passengers in the transportation system. In the agriculture field, big data shows a huge potential for solving many challenges of farming and consequently boosting the agriculture production quality and quantity. Big data analytics can be used to determine the soil quality, diseases and pest interruption,water requirement, and can predict harvesting time for crops.

There has been a critical pattern to ruminate about the utilization of massive data procedures and strategies to agribusiness as a significant opportunity for utilization of the ICT pack, for financing, and for achieving added significance inside the agriculture sector [3] [4],[5]. Applications of massive data in agriculture are not sternly regarding primary cultivation, but also assume a significant part in enhancing the effectiveness of the whole supply chain, thus reduces food security worries [4], [6]. Prospects for Big Data use in agribusiness incorporate benchmarking, IoT-based sensor network implementation and analytics, prediction models, and utilizing enhanced models to oversee crop failure risks and to lift feed efficacy in livestock farming. Thus, big data technology is to offer prescient insights to upcoming farming outcomes, enables real-time effective decision making, and modernize business measures for rapid, state-of-the-art actions, and game-changing business models [7]. Big data is predicted to modify both the degree and the organization of agriculture.

Key domains of progress for precision agriculture such as real-time forecasting, tracing of agri-food products, and remodeling of business practices [8]. More extensive big data application is probably going to transform both farm organizations and the more extensive supply chain in unfamiliar ways.

## 3.2 Importance of Big Data

Advancements in computing technology make large data available to Government, which can cost effectively utilize it in delivering social services and bring about improvement in social development indices. However the real revolution is not machines that calculated data but in data itself, and how we use it (Mayer-Schönberger, 2013).
Big data can be used for organizations in the following way:

1. Cost Saving
2. Availability of data on time
3. Protect online reputation of the organization
4. Know your customers better.

Government organizations in countries across the world utilize the Big data services to monitor the health and safety of its citizens at the same time maintain a vibrant economy that could bridge the difference between haves and have not's. Big data if properly utilized brings astonishing positive outcomes for public administration in terms of its efficiency and overall client satisfaction. Digitization of information and related data flexibility, progress in artificial intelligence & automation all these lead to novel and powerful methods for accessing and using information in new ways hitherto not available.

# Chapter 4

# Machine Learning Techniques in Agriculture

There is a broad literature on different machine learning algorithms that have been employed in diverse application areas in agriculture. Identifying the ideal method for guaranteeing accuracy and constancy for a specific application in agriculture is significant. SVRs demonstrated robustness with outliers and noise presence with better estimation accuracy upon comparison with ANN [9]. ANN and SVRs when used for mapping of soil organic stocks (SOC) produced comparable performance. Several regression models were evaluated to find appropriate techniques that realize great accuracy and better generality for yield prediction abilities. Neural networks, despite their site-dependency, ascertained robustly, however, the SVR model employed was highly accurate though being fast computationally [10]. ANNs, RFs, and SVMs have mostly been testified as classifiers, yielding great accuracies [11], [12]. Deep learning techniques are the utmost promising models for segmentation applications of agriculture image data sets. Finally, in Table 3 we discuss the reliability issues, computational characteristics, and threads of analysis of the models explained.

Graphical models are not apprehensive towards inputoutput pattern modeling; however, they model autocorrelation between the input parameters (variables) [13]. There exist several variants of GMs that model data based on input-output interdependencies, for instance, conditional random fields (CRFs) [14], [15]. CRFs take up probabilistic systems unlike NNs and SVMs; i.e., they form inputout variable relations by a probability distribution $P(Y \mid X)$, instead of a definite capacity of the configuration $Y = f(X)$.

Though, having a similar fundamental principle; the models learn the $P(Y \mid X)$, upon provided with training data $(X_i, Y_i)$, $i = 1, 2,.., n$ by optimizing a proper loss function. Upon learning the best probability distribution from the training data, inference can be performed to estimate the target output. In agriculture, GMs show a substantial potential for modeling big data applications such as spatial disease distribution [16], operational decisions, modeling traits, etc.

The most specific challenges that occur with ML in precision agriculture are variable spatial-temporal resolutions and missing due to several reasons like IoT device malfunctioning, communication failure, bad weather prevented remote sensing image acquisition, etc. It is in this way important to have AI models that can adapt to missing information. All the recent ML and DL models designed for plant disease and

| TECHNIQUE | COMPLEXITY | CHARACTERISTICS & LIMITATIONS | CURRENT APPLICATIONS |
|---|---|---|---|
| ANN [36-38] | High | High power consumption;<br>Susceptible to overfitting;<br>Require large datasets;<br>Parameter tuning procedure is time-consuming | Pattern classification and attribute mapping;<br>Crop estimation;<br>Soil properties analysis and estimation |
| DT's [36] | Low | Sensitive to trivial variations in training data;<br>Unstable occasionally;<br>Susceptible to overfitting | Classification;<br>Crop yield from soil variables |
| SVM'S [39][40] | Low | Fast and accurate;<br>Easy to implement;<br>Require small training samples;<br>Robust to noise in training data;<br>Mostly used for classification;<br>Less susceptible to overfitting | Classification of diseases;<br>Soil mapping;<br>Retrieval of vegetation attributes;<br>Estimations |
| RFs [43] [44] | High | Overfitting problem;<br>High efficiency;<br>High prediction performance;<br>Small training time;<br>Easy parametrization | Attribute mapping;<br>Classification and regression |
| DLs [46-48] | High | Highly promising for agriculture applications;<br>High accuracy;<br>Require large datasets;<br>High computational cost | Classification and regression applications;<br>Measure features;<br>Yield estimation, etc. |

ANN = Artificial Neural Networks, DT = Decision trees
SVM = Support vector machines, DL = deep learning, RF= random forest

**Figure 4.0.1:** Comparision in Machine Learning Techniques.

pest detection are not suitable for the early detection of diseases and pests, thus unable to prevent the crops from early disease and pest attacks. Thus, deep learning models for the early classification of plant diseases and pests are important.

DL and CNNs have been progressively more employed in agriculture remote sensing applications. CNN needs a hefty volume of data towards generating hierarchical features to make available semantic statistics at the output. With the expanding access to enormous quantities of aerial images from unmanned aerial vehicles (UAVs) [17] and satellites, CNNs can assume a significant part in the analyses of all this information to extract significant information. Though, the UAV-based technological adaption by farmers for specialty crops is very low [18]. There are two facades for low adaption such as preprocessing and analysis of data, as it is very complex and time-consuming to produce precise and suitable information and the inability of available commercial tools to create enough useful information from the data for specialty crops.

As UAVs can accumulate a large and unstructured quantity of data, big data-based tools (analytics tools) and cloud computing has the potential to enhance the data processing efficacy, offer high data security, and scalability, and minimize cost. Applications based on cloud computing act as a potential solution having low upfront cost, proficient utilization of computational resources, and service costs [19]. UAVs with big data analytics methods such as CNNs can be used to detect tree characteristics (height, tree health, species, canopy area, etc.), leaf disease, crop estimation, etc.

Soil quality performs a significant part in influencing how healthy plants develop. Indeed, different types of plants grow in different soil conditions or types. Understanding the diverse characteristics such as texture, structure, and chemistry of soil assists agriculture practitioners to choose the best quality crops to cultivate in their farms. To study these characteristics of soil, IoT, and other sensor networks along with ML-based big data techniques like clustering and classification methods to label soil data. Spark Mlib comprising of several ML algorithms and utilities including logistic regression, and naive Bayes in classification, K-means, GMMs in logistics. Likewise, distributed parallel association rule mining techniques can be used to determine the growth of plants.

## 4.1 Steps in Conceptual Framework

The objective of the research is to propose a framework for the application of data analytics in enhancing rural development by supporting agriculture sector. The research aims to provide a conceptual analytics platform which will help various stakeholders related to the above mentioned sectors with adequate information and forecasts; thus increase the productivity and is beneficial for socio-economic and community development.

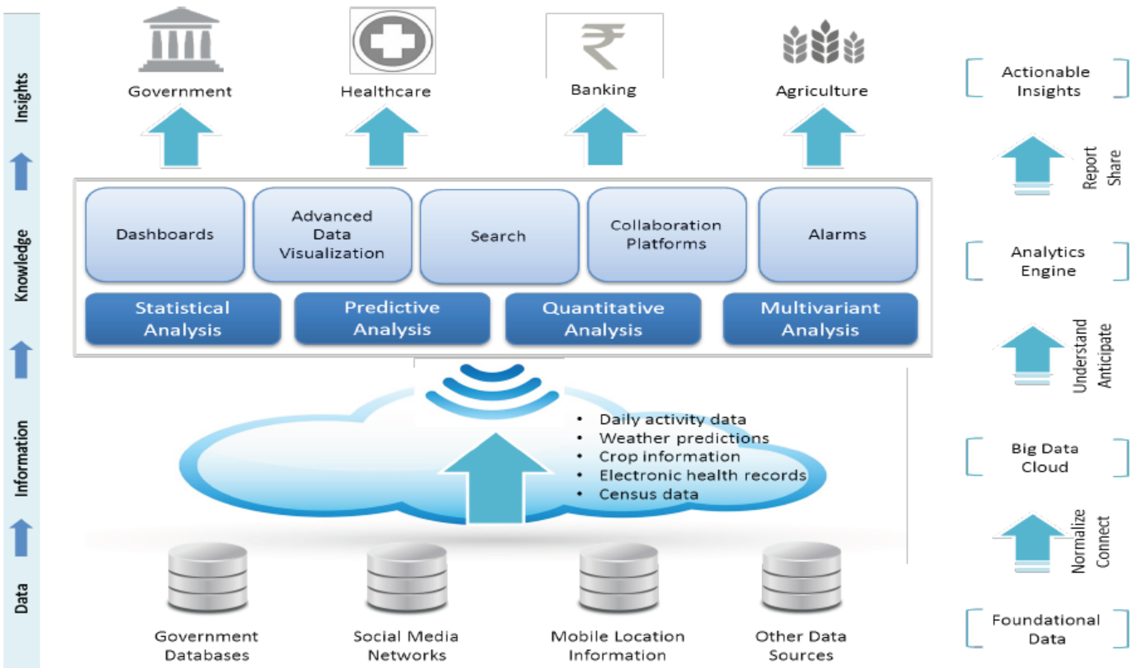


**Figure 4.1.1:** Conceptual Framework

## 4.2 Data Integration

Whether reforming an individual sector or undertaking a city wide operational reform, the first step is to integrate processes and data across diverse sector infrastructure. The ‘traditional data’ (official statistics, survey data, etc.) will continue to generate relevant information, but the digital data revolution presents a tremendous opportunity to gain richer, deeper insights into human experience that can complement the development indicators that are already collected.e-Governance would imply that beneficiaries should progressively benefit from a legally binding fully electronic and paperless exchange with rural development bodies and national administrations at all levels. e-Governance would also require applying the ‘only once’ encoding principle, allowing beneficiaries to encode data only once - those data being shared between different administrations at national and regional level - using electronic online portal functionalities. At State level, e-Governance means that digital signature should be implemented effectively by electronic portals at national and regional level and that electronic storage features are made available in national systems used for dealing with rural development projects. The data could be sourced from various applications/systems as per the availability of data across sectors. Following are the list of top data sources:

1. Media/entertainment: The media/entertainment industry moved to digital recording, production, and delivery in the past few years and is now collecting large amounts of rich content and user viewing behaviours.

2. Healthcare: The healthcare industry is quickly moving to Electronic Medical Records (EMR) and images, which it wants to use for short-term public health monitoring and long-term epidemiological research programs.

3. Video surveillance: Video surveillance is still transitioning from CCTV to IPTV cameras and recording systems that organizations want to analyse for behavioural patterns (security and service enhancement).

4. Transportation, logistics, retail, utilities, and telecommunications: Sensor data is being generated at an accelerating rate from fleet GPS transceivers, RFID tag readers, smart meters, and cell phones (call and usage data records); that data is used to optimize operations and drive operational BI to realize immediate business opportunities. The complexity of handling this expanded universe of data sources is compounded by the need to link, match and transform data across business heterogeneous entities and systems, while managing scale and timeliness. Consumers are increasingly active participants in a self-service marketplace that not only records the use of affinity cards but can increasingly be combined with social networks and location-based metadata, which creates a gold mine of actionable consumer data for retailers, distributors, and manufacturers of consumer packaged goods. The key to effective data-driven decision-making is the ability to sift through large

| Healthcare | Agriculture | Governance |
|---|---|---|
| Electronic Health Records | Kissan SMS portal system | UID card |
| Vaccination records | Community Information Centres | PAN card |
| Clinical Information Systems | AGMARKNET | Passport |
| | e-Choupal | License |
| | Agriwatch.com | Census data |
| | | Tax & expenditure info |

**Figure 4.2.1:** Data Sources

## 4.3 Data Normalization

Part of the need for new technologies for big data (versus older, legacy Relational Database Management Systems) has to do with the format of the data coming in from various applications. Data quality is a big issue. A more dynamic, flexible database schema is needed to handle the structured, semi-structured, and unstructured data that comprises today's big data. These database schema is need to be able to deal with a wide range of data formats since data formats density varies significantly across industries as shown below and Big data for development generally share some or all of these features:

1. Digitally generated: data are created digitally (as opposed to being digitised manually), and can be stored using a series of ones and zeroes, and thus can be manipulated by computers.

2. Automatically collected:system automatically extracts and stores the relevant data as it is generated.

3. Geographically or temporally traceable: e.g. mobile phone location data or call duration time.

4. Continuously analysed: information is relevant and can be analysed in real-time. Since there is no 'perfect' data, a concept of 'fit for use' is applied keeping in mind the purpose of data. The potential of big data for development is best realised when its limitations, biases, and ultimately features, are adequately understood and taken into account when interpreting the data. Data must be able to comply with all the laws international, federal and state regulations, fiscal and monetary reporting statutes and all applicable civil rights laws, including privacy and security.

| Industry | Video | Image | Audio | Text/Numbers |
|---|---|---|---|---|
| **Banking** | Medium | Medium | Medium | High |
| **Insurance** | Low | Medium | Low | High |
| **Health care** | Low | High | Low | High |
| **Government** | High | Medium | High | High |
| **Education** | High | Medium | High | Medium |
| **Agriculture** | Low | Medium | Low | High |

**Figure 4.3.1:** Data Formats and quality provided by various industries.

## 4.4 Data Storage

The cloud can offer a centralized knowledge bank which can be used to store all the industry related information. This information bank will be available to the stakeholders and other users from the sector at any place and at any time at a very reasonable cost. Vertical sectors such as financial services and retail are leading the adoption of mobile and cloud technologies in India.

Cloud-enabled device management helps in user authentication and secure file sharing and syncing of diversified application access. Real-time usage inputs data analytics to help in identifying trends, gauging utility ratios and making fact-supported decisions. These enable merchants to access all operation-related information and functionality through a Point-Of-Sale (POS) application installed on a consumer's device, such as a PC or mobile, without having to invest in specialized POS hardware, servers or storage. All the product and customer data, management, reporting, and analytics are done in the cloud. Some solutions can also work on PCs and laptops, but tablets and smartphones are the main target devices for this type of solution. e.g. Farmers in Sahara, South Africa use a cloud-based trading system that disseminates information about planting schedules, crop status, harvesting times, and market prices through mobile phones. The cloud based storage has helped health care providers with access to health record of patients remotely through a mobile device.

## 4.5 Metrics and Analytics

The first step is to connect data collected from a variety of sources: network and non-network, structured and unstructured. Unique data elements (metrics) for the subject areas viz. healthcare, governance, banking, agriculture should be identified. Analysis should be done on the overall availability and quality aspects of the identified fields. A common framework for information processing is defined to discover patterns and trends in the data. A common reference guide is required that will help tag data with following parameters in Figure

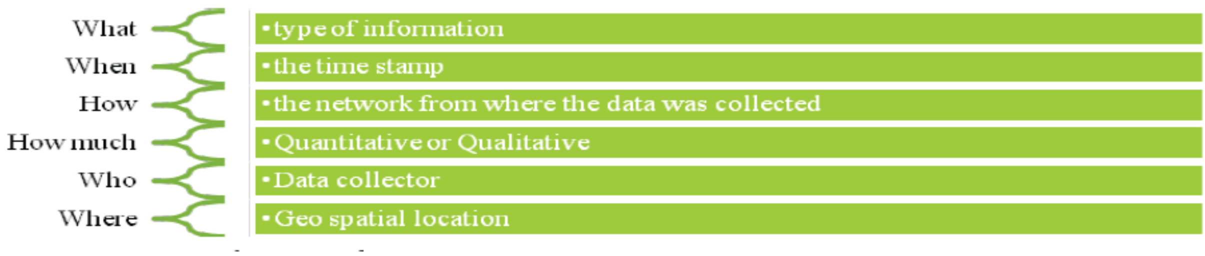


**Figure 4.5.1:** Parameters for tagging data.

Big data analytics refers to tools and methodologies that aim to transform massive quantities of raw data into 'data about the data' - for analytical purposes. Such exceedingly large data volumes cannot be analysed with ordinary assessment methods, such as sampling or simple spreadsheets. They typically rely on advanced visualization techniques and powerful algorithms. Algorithms are able to detect patterns, trends, and correlations over various time horizons in the data and can help in detecting anomalies in the form of large deviations from the expected trends or relations in the data. Visualisation tools can provide new perspectives on findings that would otherwise be difficult to realise.Biggest power of data analytics is the predictive capability that can help to determine reliable patterns and forecast what might happen in the future.

With improved real-time connectivity and data management arise the possibility to create tailored data sets, readily available for analysis. This would be the core ingredient in data-driven efficiency improvements in Agriculture. Turning big data into actionable information requires using computational techniques to unveil trends and patterns within and between these extremely large socioeconomic datasets.It is important to shift from analytics application silos to more generic, horizontal analytics environments that take in a wide array of data sources, while supporting a variety of applications and services.

Cutting-edge IT components like data storage, data management and network resources can work in harmony with domain- specific analytic logic (data models, rules sets, and so on) to bring benefits in terms of agility and scalability. e.g. Indian Agricultural Statistical Research Institute has developed a Decision Support System on nutrient management in crops. Under the e-Governance programme, Soil Health Card software has been standardized and in collaboration with Indian Institute of Soil Science, Bhopal, web-based software has been developed to provide integrated nutrient management recommendations using ‘Soil Test Crop Response’ method for 8 states.

Analytics to be fully effective can be automated and run at regularly scheduled intervals, which enables immediate identification of potential high-risk exception transactions in real-time.Furthermore, informed and effective evidence-based decision-making in business and society would be facilitated by access to insights based on analysis and interpretation of more accurate and up-to date data. Leveraging big data can reduce time lag and human inputs/errors in data collection, production and transmission.

# Chapter 5

# Operation Cycle

The above discussion on the existing work on smart agriculture and the potential of integrating evolving technologies namely AI and big data to bring revolutionary changes, benefits and solve many problems of sustainable agriculture. In the technologically advanced big industrial farms, field management looks different from the traditional farms following the operating cycle represented in figure 6.0.1.

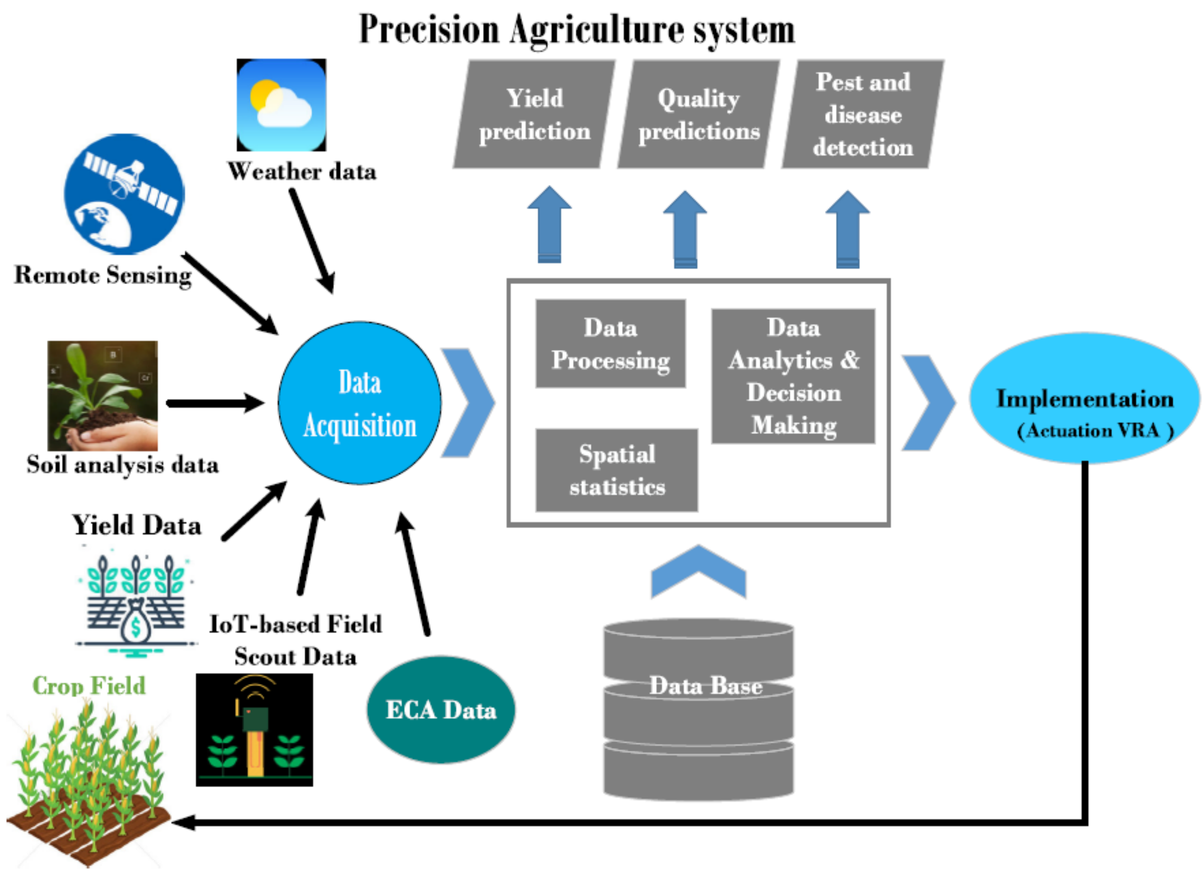


**Figure 5.0.1:** Big data-based precision agriculture system representation.

Different IoT sensors apart from traditional sensors are used for crop monitoring and collecting required important data from them. These sensor devices can be directly deployed in the fields, agriculture robots, autonomous platforms, machines, or weather stations. Different parameters can be calculated in real-time by using IoT sensor networks enabled with a high-speed data network. Remote sensing from artificial satellites has performed a vital part in the development of precision farming by making field data remotely accessible. American Landsat satellites, the European Sentinel 2 system, RapidEye constellation satellite system, GeoEye-1 system, and WorldView-3 are the important satellites supplying agriculture data in the form of multispectral data, multispectral RGB imagery, RGB and NIR data, etc.

The application of unmanned aerial vehicles (Drones and Remotely-piloted Aircrafts) in agriculture production is gradually increasing as a measure of an efficient method to sustainable agricultural management allowing growers, agri-engineers, and agronomists to assist simplify their procedures, utilizing robust information analytics to achieve valuable insights into their crops. Drones have made careful crop monitoring easier over large areas of agricultural lands, in identifying suitable crop recommendations, the emergence of plant and population, as more precise data can assist in decisions regarding replanting, pruning, and thinning activities, and yield estimation. UAV's) [17] are very useful but still face certain challenges such as carry a limited payload, limit the use of sensors onboard, challenging data and image post-processing, vegetation shadowing during gathering imagery data, etc.

In proximal sensing, the ground platforms such as Unmanned ground vehicles (UGV) and robots that operate close to crops increase the accuracy of acquired data and one or two high resolutions of samples per unit area are reasonable [20]. With UGVs applications requesting real-time data like weed detection and removal, selective pesticide spraying, soil analysis, pest control, and crop scouting are possible. Scouting robots are used for performing specific tasks such as robot Oz (mechanical weeding), GUSS autonomous sprayer, RowBot system (fertilizing, mapping, seeding, etc.), VineRobot (vineyard management), etc. Researchers and industries are working on different projects to converge UAV and UGV for better sustainable development.

The application of different wireless data collection technologies has created massive data in agriculture. But the huge quantity of data poses a significant challenge to manage, as important information may be imperceptible by noise. Presenting information in a coherent shape is vital for end-users to comprehend the different processes in the field [21]. Mapping is the most useful technique to express spatial developments and homogeneous sub-fields from the agriculture data. Maps assist in creating management zones with interesting parameters for the efficient application of custom-made field practices for each sub-field zones. Kriging is a commonly used interpolation method to get manageable size sub-fields. Considering the enormous quantity of data generated by smart agriculture, there is several software applications employed to handle interpolation. Local Tangent Plane (LTP) featuring Euclidean geometry enables a user to establish origins and utilizes intuitive coordinate set up east-north. The systematic quantization of the LTP coordinates is allowed by grids in maps in the efficient management of agriculture production data enables the data sharing among succeeding seasons, and different field parameters of a management zone.

Software-based farm management solutions for instance Geographic Information Systems (GIS) encourage the automation of data collection and analysis, supervising, planning, record keeping, decision making, and farm operation management. These tools also help in basic tasks for record-keeping such as produce harvests and yields, scheduling farm tasks, profits-losses, tracking of soil nutrients, weather prediction, and mapping of the field, and other complicated functions for automating field management.

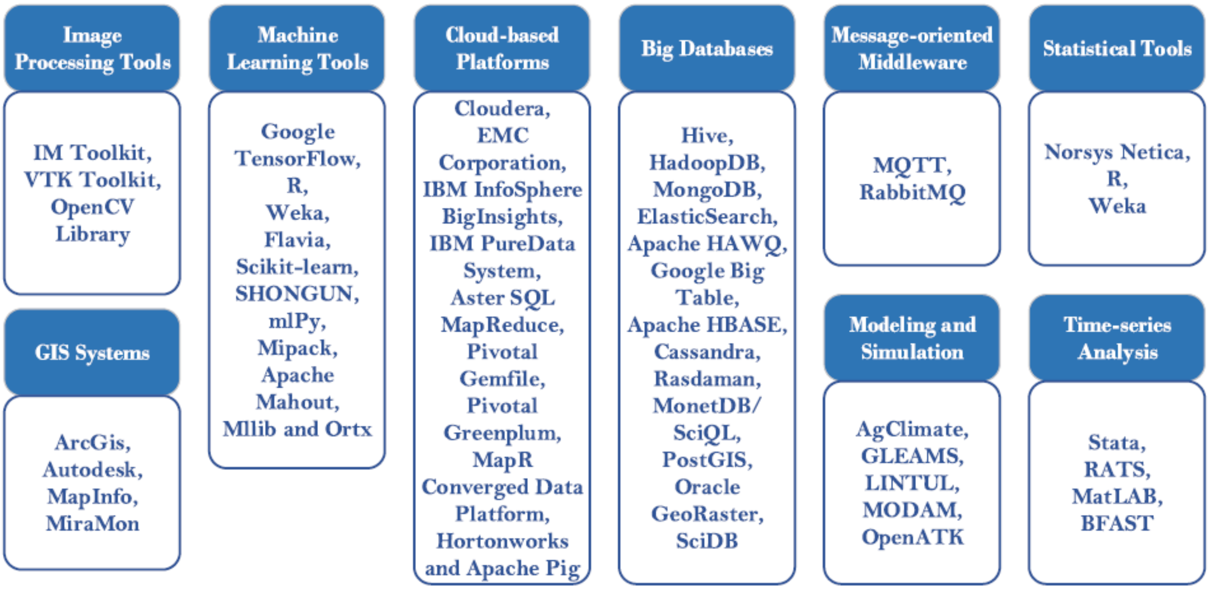


**Figure 5.0.2:** Softwares used in data analysis.

A specific GIS data management system known as farm management information system (FMIS) was developed for different applications of precision agriculture. FMIS helps growers with several tasks like operational planning, record keeping, implementation, and evaluation of executed fieldwork. Its objective is to decrease costs of production, comply with farming standards, great product quality, and safety maintenance, and supervising the farmers to make the best decisions. There are several commercial agriculture information management systems for instance ADAPT, WinGIS, SpiderWeb GIS, AGERmetrix, FieldView, SST software, AgVerdict Inc., Trimble, etc. addressed not only for farmers or producers, nonetheless to other participants' in the agriculture sector i.e., in the agriculture supply chain from farm to fork. However, the efficiency of recommendations from these software tools depends upon the parameters encompassed in the design of algorithms of that particular software platform. For example, DSSAT produces outputs by taking experimental data for crop model evaluations, permitting users or growers to compare the simulated results and observed results, which is significant if real-world decisions are established on modeled outcomes. There are other wide varieties of big data analysis software tools available in agriculture [22] (see Figure 6.0.2).

It is practically difficult for people to manage complex agriculture data to make better decisions due to several field parameters involved in farm management. In such scenarios, AI with DL, genetic algorithms, ML, or expert systems can assist with its reasoning, and modeling abilities can perform a vital role in precision farming, facilitating to understand of all the available data. Therefore, precision agriculture presents a huge application space for all types of core technologies in AI because agents functioning in uncontrolled situations. A fuzzy logic-based decision support system is designed in [23] for kiwi, potato, and corn with input variable parameters as rain forecast and soil moisture. Likewise, [24] developed a DSS to estimate the weekly irrigation requirement for citrus orchards by taking soil moisture and climate data; it uses a real-time soil parameter measurement control system to avoid errors. DSS is the most robust and dependable by considering several parameters, however, some processes stay controversial as different objectives can result in diverse solutions at different times depending on the need set by users or others engaged with the procedure. Numerous DSS systems have been proposed in the literature considering different use cases having different objectives. Thus, the use of DSS tools is influenced by their usability, performance, cost-viability, significance to growers, and suitability with compliance requests. The application of software tools for decision-making in precision agriculture is considered valuable as these tools enhance management efficiency than other tools. Though, there is yet far to make innovation-based tools adequately attractive, simple, intuitive, and nice for farmers to adopt. On the other hand, producers need to be trained appropriately until these technological tools can be easily managed.

Variable-rate machines are capable of executing several agriculture errands operated by automatic systems. The application of variable rate technology (VRT) for sitespecific crop management (SSCM) can improve profit and diminish the environmental impact by executing tasks precisely. Delineation technique usage in management zones can increase the efficiency of farms for instance applied delineation methods for variable rate nutrient use cases, which enhances the farm efficacy than traditional uniform-rate use cases, and impact on the environment was minimized. Different machinery manufacturing companies such as CLAAS, CEBIS MOBILE ISOBUS, etc. are developing various VRT based commercial solutions to perform different applications of precision agriculture. Variable-rate harvesting (VRH) or automatic differential harvesting is the other type of variable actuation, which tries to harvest according to initially specified management zones. Other than efficiency and usefulness, the cost is one of the vital parameters to reflect for the acceptance of these technologies. Thus, the pervasive availability of economical electronic components will favor the adoption of these digital applications throughout the world including small farm holders.

## 5.1 BIG DATA-BASED DECISION SUPPORT SYSTEM FOR CROP SELECTION

The proposed system architecture maintains the data collected at every stage of agriculture production and supply chains such as soil moisture, weather, and environment data, crop yield and harvest, demand and supply data from the supply chain, food processing data from food processing industries, and pesticides used by the farmer. Figure 4 represents the overview of the decision support system for crop selection in the proposed system. This data is at cloud storage and other local databases, are used to abstract the relevant information about the quality of soil like nutrient level and pH, analyze seed characteristics, sorting of food, the weather patterns, marketing and trade management, and the existence of food hazards by relating biotic or abiotic data with development and probabilistic existence of pathogens, pests, and toxicants. Big data analytics recommend the best appropriate crop for agriculture practitioners to select for which there will be demand. The system keeps track and maps the crops with the corresponding demand and prevents farmers from overabundance harvesting of the crops with the predicted demand.

## 5.2 CROP MANAGEMENT,GROWTH MONITORING,and PRODUCE QUALITY

The raw estimation of vital parameters from the farm data necessities to be processed efficiently to convert numbers and pictures into beneficial info. The growth and quality of the crops can be monitored from the image data collected from the farm and then applying different image processing techniques such as OpenCV, Matplotlib, Sciki-image, etc., in Python. We can measure the height and width of plants along with the quality of the crops by extracting the different features of crops based on the color of the leaves and crops from the images.

In the meantime, other IoT devices are used to collect all other environmental parameters and are stored in the cloud. IoT networks deployed in the food processing industries collect information about each step in food production from the quality of raw material used to other ingredients used to produce the final food product. All these values can be stored in the network in any file format such as CSV file format, etc. After initial analysis and determination of how indispensable the extracted features are.We can employ some of the abstracted data to train our machine learning or deep learning models. Then these trained deep learning algorithms/models are used to extract the required features from large data sets collected from the field.

The crop yield estimation targets to analyze factors that affect and influence the production, like irrigation, natural soil composition, and its physical structure and topography, climate and weather, crop stress, crop diseases, and pests, etc. It facilitates efficient management of resources; a timely and precise estimation of products can offer a reliable base to decision-makers to ascertain if there will be a scarcity or excess, therefore, to respond appropriately according to the conditions.

## 5.3 SUSTAINABLE USE OF RESOURCES

Driving advancements in the technologies, such as AI, IoT, and drones using big data in its processes to increasingly enhance sustainability in agriculture. Since the volume of arable land is not increasing, the groundwater levels are going down, and the soil quality is not increasing these technologies could make sure optimal utilization of arable land, water, and other resources to meet global requirements at the same time conserving resources for the future generations. Thus, big data have the potential to provide solutions (practical and scalable) that can assist in natural resource conservation, thus could sustain agriculture.

## 5.4 REDUCE PESTICIDE USAGE

With the implementation application of computer vision, machine learning, and robotics, agriculture practitioners can employ AI to manage weeds. From the data collected from the farms with the help of IoT devices, AI can assist in data abstraction to locate the weeds in fields and spray only to the specified locations where the weeds are. Thus, reduces the consumption of pesticide spraying an entire field and leaves less chemical on the agriculture produce relatively the amount of chemicals typically sprayed.

## 5.5 PLANT DISEASE DETECTION

Agriculture production and quality face a major threat from plant diseases. Recently, several deep learning-based neural networks have been developed to identify plant diseases but all the models work when the diseases are fully developed in the crop and thus have minimum effect on increasing the quality of the produce. In the fields, plant disease detection and monitoring at early stages are difficult, time extensive, and costly. But to increase the crop quality and yield, deep learning algorithms for timely classification and recognition of disease are required. This can be achieved by preparing plant pathology datasets with diseases in the early stages.

## 5.6 SYSTEM FOR RISK MANAGEMENT

Managing risks because of the farm location, type of soil, and mostly to heat stress or freeze involves essential significance in precision agriculture. A particular condition for cultivation is the impact of the climate and particularly its volatility. The amalgamation of different datasets is a critical process for data interpretation for this use case. Provincial climate patterns are utilized to join data from global patterns with local and provincial meteorological histories to give climate information to more modest spatial units and support real-time adoption of the environment and climate changes. For example, discussed a situation of big data employed in the forecasting of rainfall by taking benefits of meteorological big datasets. The outcomes show a considerable prospective of data fusion in precision agriculture.

## 5.7 AGRICULTURE MANAGEMENT SYSTEM

ICT empowers farmers to share data, set up collaboration, and work together. As agriculture practitioners become connected, software-defined management frameworks arise. Rural administration frameworks emerge to give accounting administrations, linking growers with farm owners and administrators, and offer benchmarking capacities to agriculturalists by associating them. Their point is to support farm administrators and agribusinesses throughout the world, incorporate, and examine a tremendous quantity of information from real-time sources to help their decisions in business. Such frameworks give smart cultivating solutions. Smart cultivating is a term that broadens precision agriculture by putting together administration errands concerning handle explicit information as well as information upgraded by setting and circumstance awareness, set off by real-time events [22]. Taking examples of studies conducted on small farms in the developing world specify that farmers are unable to trade their produce because of oversupply or inadequate information. Tools for higher productivity and require estimations can facilitate crops to be incorporated into the global supply chain [23].

# Chapter 6

# Big Data Challenges

Gathering and examining huge data produced via IoT networks and wireless sensor networks, comprising digital images and more data from UAVs, satellites, and data fusion with existing data present difficulties to the effective execution of smart farming. Arising technologies about data mining techniques and artificial intelligence techniques are potential methods to achieve intuitions from said information [22], [24], [25]. These techniques can assist with dissecting greater and more unpredictable information, uncover covered-up models, and uncover trends quickly and precisely. The capability of these strategies in massive data investigation has not been sufficiently valued in farming for various reasons inspected underneath.

The greater part of the accessible open frameworks referenced already result from the latest projects, their issue is still extensively wide embraced, to conclude the final accomplishment. Large numbers of them may in any case be underdeveloped and have not attained their maximum capacity yet. Most of these applications of big data are suited for large industrial farms (such as Monsanto) that now employ big data in process of decision making and have the infrastructure to access data, resources, and most importantly access to finance [26]. There is very little work done on small farms around the developing world. Big data has the potential to support nonindustrial farms, however moral and ethical queries regarding availability, cost, and funding need to be addressed to attain these advantages. If this trend continues, the benefits from data-driven precision agriculture will remain available to only big industrial farms.

## 6.1 Data Collection Challenges

In precision agriculture use cases, enormous data is generated from different sources. Merging data from an assortment of sources raises worries about the issue of information quality and information merging, and the access to gathered huge information raises apprehensions about safety and protection.

Data-driven techniques request uncontaminated and applicable information be used. Incomplete datasets obliterated information, and the presence of exceptions or inclinations in the training set influences model precisions. The analysis of information quality requests critical human contribution and expert knowledge. However, even semi-computerized approaches are not useful concerning huge volumes of information. The act of large information assortment likewise increases apprehensions over availability and security. The capability of academicians and researchers to perform Big Data oriented examinations firmly relies upon the accessibility of farm data. Suggestions for administering security, ownership of data, protection of data, and information use ought to be set by farmers' coalitions and farming technology suppliers.

## 6.2 Challenges in Big Data Analysis Techniques

Big data requires extraordinary methods to proficiently process a huge quantity of data with infinite running time. Hypothesis analysis and ML are the largely employed methods for data exploration. Agricultural data analysis is mostly statistical. AI techniques do not consider any pre-established relations among variables from the hypothesis however start from the data to look for potential connections between variables [27], [28].

The gathered datasets are huge and complex making it hard to manage normal AI procedures. Such methods regularly perform inadequately when applied to agricultural data. Scalable and versatile methods are expected to adapt to voluminous information. Besides, enormous information gathered in agriculture disrupt normal suppositions basic a few AI and analytic techniques, for example, the independence and identical distribution of data. Big data generated from farms shows spatial-temporal auto-correlation, has heterogeneity and high dimensionality, is non stationary, and as a rule, must be handled in a constant manner [22].

To control the datasets that go along with precision or smart farming, analytics techniques need to an extent in aligned and distributed means, high computational complication. Technological developments in cloud computing capabilities and distributed storage models can assist in this course. Distributed computing may be employed to incorporate data sources in various areas, and afterward, the data can be apportioned into an appropriated and parallel model. The integration of AI and distributed computing execution procedures gives potential approaches to deal with huge data. Established models, ought to remain compatible with distributed computing, but not all AI models are suitable for execution in distributed form. As an effective model, parallel SVM (PSVM) [29] decreases memory and time utilization and in [30] a versatile AI administration is presented for stream handling and realtime processing.

## 6.3 Availability of Computing Infrastructure

Apart from novel analytical standards for information abstraction from big data-compatible distributed frameworks and advanced wireless communication solutions are also required to implement big data in precision agriculture. Management of farms faces several challenges in executing real-time analysis and delivery of heterogeneous and multidimensional data channels from various sensor networks [31] Platforms for real-time data analysis are required to manage data collected from remote sensing online and fuse offline data with it from other distributed data sources. Since precision farming greatly depends on event monitoring this necessitates data analysis and therefore needs lesser latency and greater bandwidth. Hadoop (open source) is right for parallel processing and applications for execution of cutting-edge analytics on enormous data stored have mostly developed on Hadoop (Apache Software Foundation, 2019b). However, Hadoop is inappropriate for real-time data processing applications. Apache

Storm, Spark, and Flink are appropriate for real-time data stream processing. Several modules such as Mlib and GraphLab for providing ML operations, while Tensorflow like tools are intended to develop sophisticated1 ML models such as CNNs, and DNNs.

## 6.4 Managing Growing Data and Real-Time Scalability

An immense quantity of images and video is produced progressively through several devices during plant growth monitoring, which prompts several challenges in storing and processing all that data. Moreover, most of the data generated in agriculture are amorphous or semi-structured, not stable for storing in customary databases like MySQL, SQL Server. Management of such enormous unstructured data is considered a big challenge.

It is imperative to give consumers visual data in a real-time mode to empower them with efficient fast decision-making capabilities. This demands advanced big data platforms with real-time data handling capabilities across the network stages such as collecting, processing or analyzing, and visualizing. But real-time analyzing such an enormous amount of data is challenging near the source due to inefficient computing infrastructure.

## 6.5 Data Management Uncertainity

One disruptive aspect of massive data is the utilization of variability of advanced data management techniques whose intentions are to strengthen operational and analytical processing significantly. These methodologies are generally grouped into the NoSQL framework category that is distinguished from a traditional relational storage management system. There are various NoSQL methods. Some support a hierarchical object interpretation employing standing encoding techniques like XML, BSON, or JSON related to respectively managed data entities while others are using the key-value concept of data storage, fundamentally supporting a schema-less model. Databases based on graphs maintain the interrelated relationships between various objects. And several other standards are evolving continuously.

In fact, inside every one of these NoSQL classifications, many models are being created by several organizations both for-profit and non-profit. Each approach is fit diversely to key performance measurements-a few models give extraordinary adaptability; others are excellently scalable as far as execution while others support a more extensive scope of functionality.

To be specific, the wide assortment of NoSQL tools and designers, and the market status impart a prominent level of uncertainty to the data management landscape. Selecting a NoSQL technique is difficult, but choosing the wrong data management framework can result in enormous errors and loss if the selected NoSQL tool from a particular organization doesn't satisfy the expectations or if a different data management system is adopted by a third party for application development. Therefore, to select big data management techniques users need to consider their respective applications and performance requirements along with the mitigation of risks of the underlining technology.

# Chapter 7

# Conclusion and Recommendations

The ever-growing accessibility of information through developments in ICT appears promising for improving innovations on indispensable decision-making through enhancing precision and generalization capability of models. Besides, learning from the enormous quantity of data generated from precision agriculture practices is anticipated to create substantial opportunities and transformational perspectives for precision farming. With the advancement in big data, traditional learning methods are not naturally proficient or scalable adequately to process huge quantities of heterogeneous, multi-dimensional, and spatiotemporal data. Innovative ML techniques such as CNN, big data analysis methods present higher precision, flexibility, vigor, and performance.We have provided a comparison and discussion on the different ML techniques in precision agriculture.

Agriculture production challenges are growing, creating the necessity to comprehend the complicated agriculture environments more crucial than ever before. Several ML techniques due to their data mining capabilities from agriculture data are extensively being employed in smart farming. Several challenges facing by big data and AI in precision agriculture are classified appropriately.

Automation and application of AI, drones, IoT, robots, and big data are anticipated to perform a significant function in various agriculture areas in addition to precision farming. Employment of high-performance data-driven scalable learning methods provides better real-time decision-making capabilities and automates various agriculture processes, and thus can transform conventional farm management into artificial intelligence systems. Emerging domains of cutting edge ML and data mining converged with accessible datasets and strategy structures are required to act instrumental in addressing the challenges of agrarian production regarding sustainability, efficiency, climate change, and food security.

A data-driven system benefits every single stakeholder engaged with the agriculture business right from agriculture practitioners (farmers) to consumers, financial institutions, food processing industries, and several others. Even though the best of its capabilities are still unexplored that it has to offer for value generation, it has already begun to get enormous revolutions in the agriculture industry. Some of the various benefits AI and big data offers include the development of healthier and superior products because of the availability of the new plant genome sequencing techniques, precision agriculture methods help in inferring conversant decision making, and the utilization of IoT sensor devices and analytics techniques help in thwarting the food wastage and food-borne diseases.

# Bibliography


[1] M. Torky and A. E. Hassanein, "Integrating blockchain and the internet of things in precision agriculture: Analysis, opportunities, and challenges," *Computers and Electronics in Agriculture*, p. 105476, 2020.

[2] H. M. Kim and M. Laskowski, "Agriculture on the blockchain: Sustainable solutions for food, farmers, and financing," *Supply Chain Revolution, Barrow Books*, 2018.

[3] M. Dakshayini and B. B. Prabhu, "An effective big data and blockchain (bd-bc) based decision support model for sustainable agriculture system," in *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*, pp. 77–86, Springer, 2020.

[4] N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, and A. Martynenko, "Iot, big data and artificial intelligence in agriculture and food industry," *IEEE Internet of Things Journal*, 2020.

[5] M. N. I. Sarker, M. Wu, B. Chanthamith, S. Yusufzada, D. Li, and J. Zhang, "Big data driven smart agriculture: Pathway for sustainable development," in *2019 2nd international conference on artificial intelligence and big data (ICAIBD)*, pp. 60–65, IEEE, 2019.

[6] H. Tarik and O. J. Mohammed, "Big data analytics and artificial intelligence serving agriculture," in *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 57–65, Springer, 2019.

[7] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *Journal of Big data*, vol. 4, no. 1, pp. 1–15, 2017.

[8] S. A. Bhat and N.-F. Huang, "Big data and ai revolution in precision agriculture: Survey and challenges," *IEEE Access*, vol. 9, pp. 110209–110222, 2021.

[9] L. Pasolli, C. Notarnicola, and L. Bruzzone, "Estimating soil moisture with the support vector regression technique," *IEEE Geoscience and remote sensing letters*, vol. 8, no. 6, pp. 1080–1084, 2011.

[10] G. Ruß, "Data mining of agricultural yield data: A comparison of regression models," in *Industrial Conference on Data Mining*, pp. 24–37, Springer, 2009.

[11] P. Papajorgji and P. M. Pardalos, *Advances in modeling agricultural systems*, vol. 25. Springer Science & Business Media, 2009.

[12] I. Nitze, U. Schulthess, and H. Asche, "Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification," *Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil*, vol. 79, p. 3540, 2012.

[13] F. R. Bach and M. I. Jordan, "Discriminative training of hidden markov models for multiple pitch tracking [speech processing examples]," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5, pp. v–489, IEEE, 2005.

[14] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[15] M. Wytock and Z. Kolter, "Sparse gaussian conditional random fields: Algorithms, theory, and application to energy forecasting," in *International conference on machine learning*, pp. 1265–1273, PMLR, 2013.

[16] R. R. Vatsavai, A. Ganguly, V. Chandola, A. Stefanidis, S. Klasky, and S. Shekhar, "Spatiotemporal data mining in the era of big spatial data: algorithms and applications," in *Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data*, pp. 1–10, 2012.

[17] J. Kim, S. Kim, C. Ju, and H. I. Son, "Unmanned aerial vehicles in agriculture: A review of perspective of platform, control, and applications," *IEEE Access*, vol. 7, pp. 105100–105115, 2019.

[18] S. Ghatrehsamani, T. Wade, and Y. Ampatzidis, "The adoption of precision agriculture technologies by florida growers: A comparison of 2005 and 2018 survey data," in *XXX International Horticultural Congress IHC2018: VII Conference on Landscape and Urban Horticulture, IV Conference on 1279*, pp. 311–316, 2018.

[19] J. Varia, "Best practices in architecting cloud applications in the aws cloud," *Cloud Computing: Principles and Paradigms*, vol. 18, pp. 459–490, 2011.

[20] F. Abbas, H. Afzaal, A. A. Farooque, and S. Tang, "Crop yield prediction through proximal sensing and machine learning algorithms," *Agronomy*, vol. 10, no. 7, p. 1046, 2020.

[21] X. Cui and Z. Gao, "A standard architecture of agricultural big data for deep learning," in *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 908–911, IEEE, 2020.

[22] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Computers and Electronics in Agriculture*, vol. 143, pp. 23–37, 2017.

[23] E. Giusti and S. Marsili-Libelli, "A fuzzy decision support system for irrigation and water conservation in agriculture," *Environmental Modelling & Software*, vol. 63, pp. 73–86, 2015.

[24] H. Navarro-Hellín, J. Martinez-del Rincon, R. Domingo-Miguel, F. Soto-Valles, and R. Torres-Sánchez, "A decision support system for managing irrigation in agriculture," *Computers and Electronics in Agriculture*, vol. 124, pp. 121–131, 2016.

[25] R. H. Ip, L.-M. Ang, K. P. Seng, J. Broster, and J. Pratley, "Big data and machine learning for crop protection," *Computers and Electronics in Agriculture*, vol. 151, pp. 376–383, 2018.

[26] A. Weersink, E. Fraser, D. Pannell, E. Duncan, and S. Rotz, "Opportunities and challenges for big data in agricultural and environmental analysis," *Annual Review of Resource Economics*, vol. 10, pp. 19–37, 2018.

[27] D. P. Peters, K. M. Havstad, J. Cushing, C. Tweedie, O. Fuentes, and N. Villanueva-Rosales, "Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology," *Ecosphere*, vol. 5, no. 6, pp. 1–15, 2014.

[28] H. Zheng, L. Chen, X. Han, X. Zhao, and Y. Ma, "Classification and regression tree (cart) for analysis of soybean yield variability among fields in northeast china: The importance of phosphorus application rates under drought conditions," *Agriculture, Ecosystems & Environment*, vol. 132, no. 1-2, pp. 98–105, 2009.

[29] E. Y. Chang, "Psvm: Parallelizing support vector machines on distributed computers," in *Foundations of Large-Scale Multimedia Information Management and Retrieval*, pp. 213–230, Springer, 2011.

[30] A. Baldominos, E. Albacete, Y. Saez, and P. Isasi, "A scalable machine learning online service for big data real-time analysis," in *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD)*, pp. 1–8, IEEE, 2014.

[31] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proceedings of the IEEE*, vol. 104, no. 11, pp. 2207–2219, 2016.