Deep Science Publishing
https://doi.org/10.70593/978-81-988918-6-0_6

Chapter 6: Applying big data analytics to optimize resource use and boost farm efficiency

6.1. Introduction

The accelerated development of science and technology has raised people's living standards, but the rise in human population and corresponding increase in demand for food continues to be a major challenge. Agriculture is not only able to meet the food needs of the population, it is also important for economic development, especially in developing countries. However, ordinary farm production is limited in many aspects, such as measurement of production, varieties, cultivation techniques and management deficiency. Therefore, farm yield is often uncoupled from resource input, or even worsened. Massive amounts of agricultural data have emerged due to advanced technology and instrumentation, which present an excellent solution for improving farm productivity. Big data can be described with three characteristics: volume, velocity and variety. Data volume is growing at such rapid rates that it is becoming increasingly difficult to manage within traditional database environments (Kamilaris & Prenafeta-Boldú, 2018; Liakos et al., 2018; Klerkx et al., 2019). Data of many types is being generated at an increasing velocity. Various cloud services and databases have emerged and been utilized to address a plethora of typical, individual problems. Data contains variety, which reflects in its heterogeneous nature, such as open data, remote sensing data, and cloud data. Analysis of Big Data offers the potential to explore hidden patterns in data in revealing new knowledge and using new technology to improve applications. Big Data generally describes data that moves quickly, comes in deep volume, and is available in a wide variety of forms, from raw data to curated data products, and can provide novel insights when correctly analyzed and non-trivially processed. Understanding or optimizing such systems is now one of the grand challenges that science and industry face, yet are circumventing specialized tools or statistical induction. Data science, at the intersection of mathematics, statistics, computer science, and domain area expertise, has emerged as a new field. Big Data are characterized by great volume, velocity, variety, and low value density (Verdouw et al., 2021; Zhang et al., 2021).

6.1.1. Overview of Agricultural Data Landscape

The agricultural sector worldwide is experiencing a paradigm shift, with increasing pressure on food production to meet demands to feed a growing population. In addition, there is a corresponding increase in demand for eco-friendly farming practices to address climate change. In the quest for sustainable agricultural growth while ensuring sufficient food supply, farmers are resorting to the use of various sophisticated tools and technologies. Use of Hi-Tech precision farming solutions incorporating IoTs, drones, sensors, etc. enable increased production via better crop management, pest management, soil management, inventory management, and risk management at all levels from micro to macro. Such technologies are capable of generating and collecting multitudes of data driven from multiple sources at various levels of operations.

The deployed techniques generate both structured data, e.g., satellite images, meteorological data, crop yield data, etc., as well as unstructured data, e.g., images, videos, blogs, reports, etc., serving as critical parameters for timely predictive and decision-analysis in various fields of agriculture. The sheer volume and variety of available agricultural data sets are creating new opportunities, as well as challenges in the realm of agriculture. Advanced analytics techniques of big data analytics provide new capabilities in terms of efficient data storage, easy data retrieval and access, new data-driven predictive techniques, faster computational capabilities, and enhanced visualization and interpretation methods, providing appropriate insights from the vast data sets. The potential applications of big data analytics in agriculture are diverse and all-encompassing, ranging from crop yield prediction to land use prediction to climate impact analysis to pest infestation prediction to price volatility to supply chain management and to much more.

6.2. Understanding Big Data in Agriculture

Big Data in Agriculture refers to the use of Big Data in agriculture, and agriculture is the largest profession in the world. Nowadays, an enormous amount of information and knowledge are stored in various databases, big-data centers, and agricultural data networks in farmers' organizations, agricultural research organizations, and government agencies. However, only a small part of this big-data pool has been exploited through big-data analyses to improve the accuracy of agricultural meta-databases, such as soil, weather, and crop growth databases; innovation of agricultural technologies; and the designs of agricultural policies and organizations. Therefore, there is increasing awareness of the importance of Big Data in studies of agricultural science, technology, and policy.

These big-data applications for agriculture are various based on the by nature sciences theoretical basement, which include life science theoretical basement, various scientific theoretical basement, socio-economic science theoretical basement, et al. The Big Data in agriculture is the new scientific paradigm, which is beyond the disciplines. Agricultural scientists may work on any stage of the Big Data process, from data creation to data analysis. So far, most scientists focus on the Big Data analytics, especially the development and use of new methods in statistical analysis and machine learning, while there are still less works conducted on the data management. The Big Data ideas, methods, and models have the potential to boost innovation in agriculture and agricultural economics in a wide range of process. However, the actualization of the potential for Big Data in agriculture depends critically on the quality and quality of updated big-data tools. Thus, in this report, we first present insights into big data agriculture ecosystem. And we sharply value and evaluate the tools as well as services related to big-data analyses.

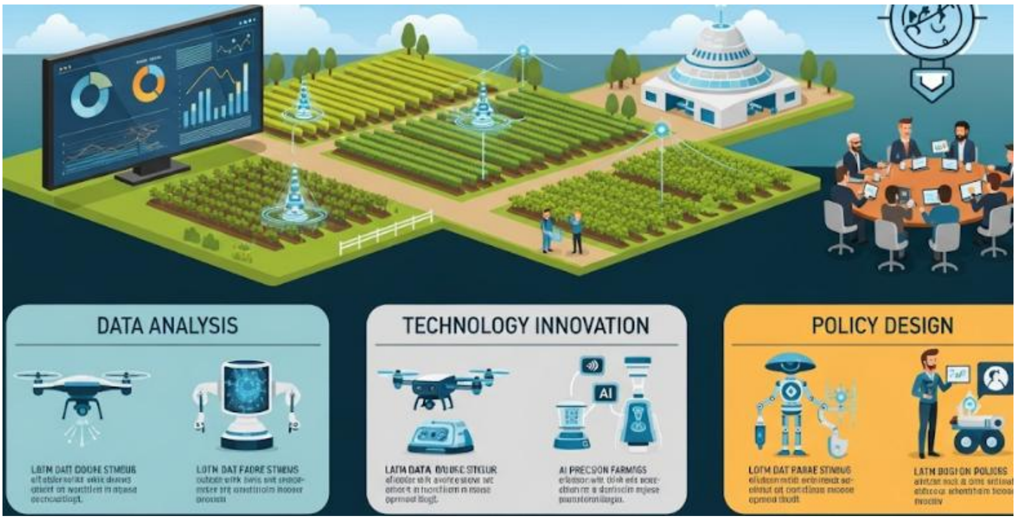

Fig 6 . 1 : Agri-Data Revolution

6.2.1. Insights into the Agricultural Data Ecosystem

Big Data is a term widely used these days, and while people may be familiar with the term, they may not fully understand what Big Data stands for and why it is important. Big Data involves large amounts of data, high-speed processing, insightful analysis, and informed decision-making to predict future outcomes. Big Data analytics is widely used in several sectors including healthcare, big retail, e-commerce, IT, and finance. Some consumer-centered companies are using Big Data analytics to predict consumer behavior trends to innovate and position themselves better. However, few of these techniques are used in agriculture. Farmers and agriculturalists generally use very little data compared to other sectors. But this gap is being slowly addressed with the advent of new-age precision agriculture concepts and tools, farmer community tools, and agricultural advisor tools. Several new private companies providing products and services using modern technologies such as drones, satellites, and AI/ML have sprung up in agriculture. In addition, new government and private funding initiatives have been launched to fund new age agricultural projects. These crowdsourcing management initiatives will pave the way for including more and more people in agriculture-centric activities for prediction analytics. However, for the successful implementation of agricultural services based on Big Data analytics, the availability and proper utilization of reliable agricultural datasets is crucial. The paper discusses the available agricultural datasets, their usability, and the challenges faced in utilizing these datasets for developing and deploying agriculture-centric services.

Big Data has permeated most aspects of human life today, and agriculture is not far behind. With the advent of new age and smart agriculture, more and more private players are venturing into the new field and offering products and services centered on monetizing sensors, satellites, drones, and A.I. and machine learning-based photo and satellite image analytics of agricultural processes such as crop growth, soil health, and pest and weed control. With these new ventures promising returns at all levels – agriculture business owners, farmers, and investors, several funds and incubators have started to invest in these companies. It has also attracted new players from outside the country who are collaborating with private players. All in all, the time is right, and the environment is conducive for Big Data analytics to take root in Indian agriculture so that farmers can help them to optimize their activities.

6.3. The Role of Data Analytics

Data analytics applies statistical, analytical, and machine learning techniques to quantitative and qualitative data to gain insights, inform decision-making, and optimize processes. To support productivity and resource conservation, agronomic decisions must be informed by spatial and temporal data integrated from farm and external sources. Modern agriculture is immersed in a tsunami of data assisted by advanced digital technologies that sense and measure agricultural processes and properties to support the transition to digital agriculture. Sensor and drone interventions have created a deluge of high-frequency and high-dimensional vectors of agricultural and environmental data measured continuously, spatially, and temporally, while Big Data technologies have developed new computational and storage capabilities to process and manage these large and growing data with unprecedented speed and efficiency, providing new opportunities to support real-time precision agriculture and inform resource-driven digitally-enabled decisions to raise productivity and conserve resources.

For example, crop yield, nutrient and water uptake, evapotranspiration, soil moisture, temperature, and weather can be measured in real-time or with high rates by remote and handheld sensing to support accurate simulations of crop growth, nutrient and water uptake, and yield for real-time farming decisions. Careful decisions in farming plans in the areas of tillage, planting, irrigation, pest and disease, fertilization, and harvesting can lead to increased agricultural income and sustainable use of natural resources including water and energy while preventing land degradation. Digital agriculture powered by Big Data can help prevent failures and project successful farming policies. Algorithms play a critical and sophisticated role in data analytics to increase the power of what is learned from the data. The choice of a proper algorithm will depend on predictions given the nature and challenge of the problem, the metric of forecasting accuracy, the feasibility of implementation and use, and the amount of usable data.

6.3.1. Impact of Data Analytics on Agricultural Productivity

There is mounting evidence that data analytics at various stages of production can boost yield and farm productivity. In particular, analytics that leverage advanced info-communications technologies such as mobile phone networks, satellite imagery, and the internet have demonstrated their effectiveness in transforming agricultural productivity. This section will discuss the need for analytics to be integrated not just at production levels, but across the whole agribusiness value chain – including supply chain, distribution, storage, trade, and retail. Such a systems approach to agricultural analytics will realize the goal of resource optimization at every stage of agricultural activity, help commercialize agriculture, and catalyze the growth of the rural economy.

It is useful to situate discussions of analytics in the Indian context. Agriculture is the primary economic sector in India, engaging 54 percent of the total labor force and contributing 15 percent of the national GDP. Yet, it has not escaped the ruler's curse of declining productivity. Consequently, increasing agricultural productivity still remains the single biggest challenge for India today. There is no doubt that enhancing productivity is a powerful instrument for generating economic growth. The increasing application of data analytics in the area of agriculture presents exciting opportunities for transforming farming through higher productivity and increased profitability. Such transformation is necessary as China and the West have established a new scientific model of growth. According to the new growth model, agriculture must play a critical role as a stockholder of growth rather than as a sector that absorbs surplus labor, and data-driven information is the invisible hand that allows agriculture to function as a high-performance engine of growth.

6.4. Types of Data in Agriculture

Accurately predicting crop yield is challenging because it is based on numerous primary and secondary factors, and input data is collected on a continuous basis. Various data types can be categorized into soil, weather, crop yield, and market data. Soil data consists of information that describes the physical and chemical properties of the soil. Primary soil data sources are local soil surveys that provide spatially distributed data on the characteristics of the soil. These are usually created using traditional sampling methods. These surveys have short sampling intervals and large prediction errors. However, with the advent of precision agriculture, different soil attributes, such as soil type, moisture, and temperature, can be measured at high sampling densities using in-situ or remote sensing approaches.

Weather data comprise information that describes climate variables such as temperature, precipitation, sunlight, humidity, and wind. Weather exerts its influence on crops through the three growth phases of growth, flowering, and ripening. Crop-specific weather data are, therefore, necessary for linking weather effects to yield in a particular crop. Weather data can be obtained from historical weather stations, meteorological departments, or satellite-enabled applications. Crop yield data provides information about the amount of crop produced during a given growing season in a particular region. Due to the inherent problems in measuring crop yield and postharvest loss estimation, reliable and comparable crop yield estimates have traditionally been compiled only for selected years, regions, and/or crops. With the increased push toward agricultural modernization, new mechanisms are being developed to estimate crop yields using remote sensing satellites and applications aiming at near real-time yield data generation.

6.4.1. Soil Data

Big Data Analytics is defined as a scientific method for extracting, processing, and analyzing massive data to convert it into actionable information. In agriculture, data may refer to the data generated through agricultural practices by using different types of sensors such as satellite, remote sensing, UAV drones, or the data available regarding agriculture or related areas such as soil, climate, crops, etc. The combined knowledge using these types of data will be useful in understanding the relationships between the climate, soil, and crops which will be useful for predicting the crop yield and for boosting the farm economy.

Data can be of two types: 1) Real-time Streaming data, which is temporary and disappears after being used at specific times (for example, demand for certain fruits at different times of a day, different seasons, different festivals), and 2) Non-real-time static data, which is permanent and can be used multiple times at different times (crops have characteristic response depending on the type of soil and type of climate). In agriculture, both types of data can be useful, but the static non-real-time data and its relationship with the real-time temporary streaming data will be useful in knowing the cause-effect structure. For example, the demand for oranges in summer should match the availability of oranges in summer at the marketplace. Otherwise, the farmer will not be benefitted. Similarly, once the seasons pass, the corresponding static data will change next time its characteristic response will depend on the nature of the soil, the climate, and the season.

Soil data is the first and most important aspect of growth and crop yield predictions. Soil as an ecosystem comprises the physical, chemical, and biological makeup which forms the natural environment for the growth of plants and agriculture. The soil acts as a medium or substrate for plant growth and life, making essential physical, chemical, and biological processes susceptible. The soil ecosystem is composed of organic matter, nutrients, microorganisms, and minerals. Organic matter is the portion of the soil composed of decomposed residues of plants and animals in which microorganisms decompose into humus, a complex organic molecule. Organic matter is the source of food for many soil organisms, both at the microscopic level and as fauna.

6.4.2. Weather Data

Weather and climate data influence the success of maximum crop production; therefore, these data types also used for assessing the optimum resource utilization and boost the farm efficiency. All economical decisions on maximum crop production such as planting date, irrigation planning, fertilizer requirements, harvesting time, and disease management should be suited based on weather and climate. Mostly, survey or interview-based historical climatic data and contemporary weather data sets from local governmental organizations are available. Remote sensing provides an alternative to in-situ data and used in different models where different data required.

The most comprehensive climate data sets are available for long periods. Big data analytics techniques utilized some weather data offered by Remote Sensing for creating NDVI time series. Delayed time series NDVI data help the researcher to conduct research in limited water availability to assess the optimum water requirement for crops. The Landsat Normalized Difference Vegetation Index (NDVI) product is freely available, and at least 40 years' time series data can be downloaded for free. Data sets from other available satellites also used and compared, among them, Sentinel-2 and Moderate-Resolution Imaging Spectroradiometer identified for optimum crop yield estimation because of higher spatial resolution with an open source and free tool.

Data sets from other available satellites also be used and compared, among them, Sentinel-2 and Moderate-Resolution Imaging Spectroradiometer identified for optimum crop yield estimation. Accurate prediction of the NDVI growth curve is essential for many applications in plant and crop monitoring. Researcher also performed data assimilation on NDVI for wheat yield prediction and the performance result indicated that assimilation has the potential of improving yield prediction accuracy for the wheat model.

6.4.3. Crop Yield Data

Crop yield is one of the most important components of agricultural production and productivity. The estimation of crop yield provides an indication on a regional, national, or global scale of the extent of available surplus food, and thus influences food security. For farmers, the adoption of precision agriculture is motivated by three main objectives: increasing productivity, minimizing operational costs, and reducing environmental impact. Crop production data is used in varying forms and at different scales to address production-related issues, and has numerous applications in the fields of agricultural monitoring and estimates of crop agroecological traits.

Specifically, for farmers, crop yield data can be used to assess normality across production years, in order to verify occurrences of production shocks, and measure the extent of these occurrences. Crop yield data has further applications, such as assessing suitability of loan amounts and the monitoring of borrower performance; estimating losses after major production shocks; improving agro-meteorological early warning systems and disaster preparedness; estimating carbon emissions and carbon sequestered to compensate or mitigate effects of climate change; estimating the share of agro-ecosystems as a whole in job creation and economic development; and producing long-term food production projections.

Crop yield is determined by the interaction of bio-physical variables; thus, the relationship between soil and climate is quite important in determining crop yield potential. Climate change is altering the bio-physical drivers of crop production and in turn affecting food production. Changing climatic conditions, greenhouse emission-induced changes in CO2 concentration, technological improvement, land use and land cover changes, and input usage have today an unprecedented influence on the world food economy.

6.4.4. Market Data

Market data consists of existing reports and prices of marketable produce. Market reports outline supply-and-demand relations for commodities. Associations of merchants and exporters usually provide information on market prices of commodities. Crop price data from these market reports can be mined for the location-based, commodity-specific, and time-based price trend prediction. For better output prediction quality, input data can include other economic factors, such as employment of the population, money supply, foreign exchange rate, and company profits. Predictive models can be trained using historical data to understand price variation patterns better. However, predicting commodity output prices, such as agricultural crop prices, is a challenging real-world problem due to volatile price patterns.

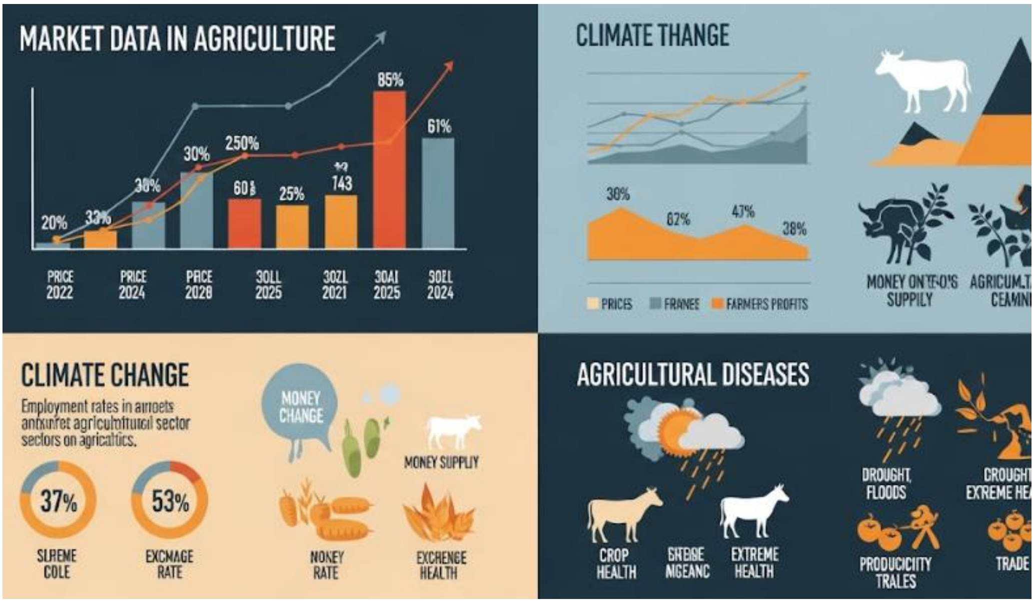

Fig 6 . 2 : The Agricultural Market Landscape

This chapter discusses the inputs provided in prices of crop and livestock, that are useful in analytics. The type of analysis is root cause analysis to find variable prices at a commodity level. Prices of different commodities from different locations behave differently. Some commodities sell at different prices at different locations. In response to high prices, more quantities of the commodity are supplied for sale. The price of the commodity is also affected by the crisis in the agricultural sector like high impact diseases in the poultry industry and climate change. The respective humidity and temperature shall be used to factor and reduce large Agriculture Price Variability. Weather forecasting and climate change for severity are the possible solutions to reduce Agriculture Price Variability. Every farmer cannot grow all the available crops that have good prices at any location. Farmers can grow those crops using soil sensors in precision farming to select crops that have the good combination of nutrients for soil at that location.

6.5. Data Collection Techniques

Farm data can be collected from multiple sources. Some of the popular data collection techniques include the use of satellite-based remote sensing, the Internet of Things (IoT)-based sensor networks, and drones. These data collection techniques provide high-resolution data at different spatial and temporal resolutions. These data sources can be used to monitor and model landscape dynamics and their influences on agricultural resources.

Remote Sensing

Satellite-based remote sensing uses part or all of the electromagnetic radiation spectrum to detect and monitor the physical characteristics of an area from a distance. Images derived from satellite-based remote sensing have been used in multiple applications, including land cover and land use change monitoring, soil property mapping, drought, flood, and disaster damage assessment, and surface-atmosphere energy exchange estimation. Multiple sensors have been launched for the purpose of agricultural data collection. The recently launched Sentinel-2 has multiple bands that are well suited for agricultural applications. The images are freely available on a weekly basis.

IoT Sensors

Sensors are devices that are used to collect data from specific locations, and these devices are getting smaller and cheaper by day. They can be mounted on a drone, vehicle, or placed in the field. The data collected by these sensors can be on soil moisture, rainfall, temperature, humidity, and crop growth. Soil moisture probes sense the volumetric water content in the soil in real time and send that information to a central server, where it can be analyzed and utilized to schedule irrigation systems efficiently. Internet of Things (IoT)-based sensor networks have changed the dynamics of real-time data collection due to sensor miniaturization, ubiquitous connectivity, and cloud storage and analytics. Using this technology, the farms can become smart, by deploying a sensor network and sending data to the cloud for visualization and action.

Drones

Drones have emerged as another key player in the precision agriculture sector. Using a drone equipped with optical, multispectral, or thermal cameras, high-resolution data can be collected quickly and at a low cost. Data collected from drones can be used in multiple applications, including crop growth monitoring, biomass estimation, frost detection, and yield estimation. Drones have been used extensively, including early-season data collection for phenotyping purposes. Data collected using drones allows dynamic screening of plant height and canopy temperature signatures of crops.

6.5.1. Remote Sensing

Remote sensing applications in agriculture are growing, leading to a need for inexpensive satellite or aerial imagery and low-cost software pipelines. Remote sensing began in earnest with Earth resource satellites. The first was launched in 1972 and has been followed by several others and other Earth-scanning missions. They have contributed an essential service by providing a unique global perspective on Earth resources and the environment and the first look at many extremes. The Moderate Resolution Imaging Spectroradiometer has collected key data sets on seasonal cycles over forested and vegetated areas and has allowed charting of important phenomena such as eddies and hurricanes.

Recent new developments include a new long-term Infrared/Visual Sensor program, airships with multi-spectral and short-wave infrared resolution for optical and infrared reconnaissance, a subcontract for an Airborne Infrared Camera, and the launch of a satellite with a sensor that provides the first operational global cloud-clearing and moisture soundings. The needs for new remote sensing satellites are to data targeting, sensor small satellite responsiveness, time- and space-multiplexed multispectral imaging, and the provision of low-cost digital image processing. The design of low-cost spaceborne digital sensors, the creation of small satellite target sensors, and the linking of these processes to a data processing unit will enable new space-based remote sensing.

6.5.2. IoT Sensors

Internet of Things (IoT) sensors are a promising area for exploring new ways to deliver real-time or near-real-time information about field conditions. They operate using ultra-low power radios and collect data on air temperature, humidity, soil temperature at different depths, soil moisture at different depths, leaf wetness, vapour pressure deficit, wind speed, rain, solar radiation, crop height, and so on. Sensor data is delivered wirelessly to a recipient where it can be stored, processed, and displayed to the user. The cost of these sensors, which is largely a result of the need for small, low-power, green batteries that can supply power over multi-year periods, has declined significantly in the past several years. It has reached the point where dozens or hundreds of sensors can be deployed across a field or a farm for a specific application or used to answer specific questions.

The number of WS data sets publicly available is growing. They are supplied by governments or commercial organizations that operate their own sensors as a service. Data sets used in publications include soil moisture and VPD data from installed across the Pacific Northwest, plastic and resin mulched soil moisture data collected from a facility in southern Ontario, and air temperature and humidity data collected at an orchard in Canada. As is made clear by the growing interest in these data sets, IoT sensors are an important tool in the big data analytics toolbox.

6.5.3. Drones

Unmanned Aerial Vehicles (UAVs) known as drones have been extensively used in agriculture as the sensors used by space systems and aircraft systems are very costly even for small areas of monitoring. Although UAVs have a limitation in flight time as compared to satellite systems but their flexible time of operation has made UAVs an efficient system in agriculture. Monitoring a very small piece of agricultural land can also be easily done. Swarms of small UAVs are flown at a very low altitude of less than 100 m from the ground to carry out different agricultural operations by reducing the time of operation drastically. UAVs, combined with remote sensing sensors such as multispectral, hyperspectral, Light Detection and Ranging and thermal sensors have proven to be useful in crop mapping and crop health and yield estimation.

With the deployment of different high-resolution cameras having a spatial resolution of about 1 cm to mapping of the small patches in the field with precision. It has been reported that the UAV mapping of small areas is more cost-effective as compared to satellite mapping. Hyperspectral and multispectral remote sensing sensors carried on UAVs have proven effective in mapping small patches of different plant species in a field. Drone mapping of the fields has become popular among the scientists as aerial images can be processed very quickly as compared to satellite images. The spectral indices computed from aerial images have proven effective in correlating with aboveground biomass and grain yield in rice and wheat crops. Research has reported the estimation of leaf area index, crop vigor index, and soil moisture index from multispectral images co-registered with LiDAR-derived digital elevation model. It has been reported that both RGB and multispectral UAV imagery can be used in assessing and monitoring crop health.

6.6. Data Processing and Analysis

As presented in the previous sections, data collection and storing of the inherent knowledge in IoT based framework has been well defined. However, as earlier stated, it integrates heterogeneous sources of data in massive volumes, is characterized by inconsistency, missing data and noise. Thus, data processing and subsequent analysis are critical in deriving useful knowledge for guiding decision making that improves overall farm efficiency and productivity. Data processing can be categorized into three main parts: data cleaning and preprocessing; data integration schemas and procedures; and data analysis techniques.

1. Data Cleaning

Data cleaning is the first step in the data processing and is an essential aspect for data quality assurance, especially for Big Data in which faults arise from diverse domains and various sources, such as sensors, cameras, weather data, farm management systems, and so on. The major challenges in data cleaning involved in data for crop production from IoT-based automated systems are the following: (1) sensors errors, malfunctions, or external environmental disruptions, (2) incomplete data due to sensor disconnections, storage media failure, and maintenance scheduled down time; (3) noisy data, defined as the forces that have the potential to invalidate the information about the currents state due to adverse environmental conditions; and (4) outlier values due to malfunctioned sensor readings, incomplete knowledge about sensor range, and unusual external conditions not found in historical databases. Thus, the removal of all such inconsistencies requires applying various available techniques to each issue or challenge accordingly.

2. Data Integration

In farming production systems, numerous factors influence the models that affect the outcome. Knowledge integration techniques that combine data and models from different domains hold great potential to produce results of expanded value. Various data integration procedures are available, both grounded in deterministic as well as stochastic theories of data fusion. Recent frameworks using sequential Monte Carlo techniques have been able to handle efficiently, data-redundant scenarios. Various sensor data fusion techniques were applied for multiple applications in crop production systems, paving the way towards effectively engaging Big Data schemes for resource optimization and system automation. Although data integration has been mentioned as an established field of inquiry, it has been primarily developed in the context of enterprise systems.

6.6.1. Data Cleaning

On most occasions, raw data is riddled with errors. For instance, errors may have been introduced due to sensor malfunction or incorrect calibration issues, which can lead to poor temperature accuracy or bad GPS data. Similarly, label or value corruption presents risk to data quality, impacting analysis results severely. An example of such a label corruption would be an error in the precipitation label during sunny weather wherein instead of “0 mm” the respective label is incorrectly logged as “25.5 mm”. Other occasional problems include large temperature deltas for the same sensor within a short time period, unreasonable or unlikely combinations of temperature or humidity measurements, humidity and temperature values that do not match the respective ranges, incorrectly reported dates and time labels, missing records or even entire months, wrong GPS coordinates, or unreasonable amount of precipitation for a specific location. In addition to errors introduced by malfunctioning or incorrect calibrated sensors, human error can also diminish data integrity. For example, human logging errors, missing values, or incorrectly stated locations can severely impact the integrity of meteorological datasets.

Prior to any analysis, datasets should be checked for such or similar inconsistencies and corrected based on domain knowledge where feasible. Potential repair methods include value replacement for missing data, delta checks for detected outliers, global threshold checks for invalid data, plausible value checks for range violations, and checks against friends. A combination of these procedures discovered large-scale errors, while local issues were identified by manual inspection. Apart from these common techniques, machine learning algorithms could also be applied to clean erroneous data. For instance, a model trained on data containing high quality data labels and input variables similar to those used could serve as a possible solution.

6.6.2. Data Integration

The objective of data integration is to bring together data residing in different sources and providing users with a unified view of them. Data integration has been around for decades and has become a critical part of an enterprise architecture. Organizations want to bring together all the data they have about their day-to-day activities and make that data available for business users to gain a full picture of the organization. They want to be able to assemble data from disparate sources at intervals of minutes or hours so business users can be alerted when things are happening in time to take action. Data warehouses and data marts have traditionally been the core of business analytics and reporting but organizations are becoming more interested in new big data sources. However, big data sources are different from traditional sources in several key ways. Big data sources can include both structured and unstructured data, data sources can be both internal and external to the organization, batch and real-time data sources can be connected, big data sources can be much larger, data can be stored in many different formats, and data can be organized in many different structures. For these reasons, the technology that has become available for data processing and integration for big data sources differs in important ways from the technology traditionally used to perform these functions for data warehouses and data marts. Commodity and grid computing: Virtually all new approaches to data processing for big data leverage commodity and grid computing. What this means is that organizations can now use clusters of inexpensive servers or services in the cloud instead of a few big box mainframes and SMP servers. The economics of these new approaches hinge on the fact that most big data processing applications are able to run in parallel and do not require a lot of interprocess communication. Organizations can now use these low-cost options not just for the batch processing of data that have traditionally been used but also are using these approaches in new ways for performance-sensitive data integration and data quality.

6.6.3. Predictive Analytics

Decision support is an essential element of modern agriculture, but one of the hardest and most important problems of this area is to make the right decision at the right time. In recent years, several decision support tools have been proposed to help farmers make better decisions. These tools include algorithms for predicting the future performance of a crop, farm field, or production system. These prediction or forecasting algorithms are also known as Predictive Analytics, where the term Predictive Analytics indicates a set of mathematical and statistical techniques and processes to produce a set of prediction or forecasting results. Predictive Analytics helps answer the following types of questions: What will happen in the future with my farm fields? When will it happen? What are the chances that it will happen? Predictive Analytics should also be a component of all analytic software packages for agriculture. Many predictive analytics algorithms have been developed to help farmers predict what is going to happen in the future. A good predictive analytics tool should provide daily predictions for the data being monitored and alert farmers when the device failure is going to happen soon.

Predictive Analytics is useful for growers because it has the potential to improve yield and quality grading, leading to sustainable development of greenhouse vegetable production. It can also provide valuable information for other applications, such as energy supply and climate-based warnings. Predictive analytics is an advanced level of analysis, which uses statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. Predictive analytics identifies patterns in historical data and builds a statistical model to replicate these patterns – allowing you to evaluate future behavior and outcomes. Predictive Analytics helps answer the following types of questions: What is going to happen in the future? When is it going to happen? What are the chances that it will happen?

6.7. Optimizing Resource Use

1. Water Management

Big data analyses of weather, climate, soil and crop traits allow efficient irrigation water use through precision decisions on when, where, and how much to irrigate, which eventually increases farm productivity and reduces producer risk. At the field level, elaborated near real-time satellite imagery monitoring systems help optimize irrigation schedules based on current field conditions in order to avoid any water-related crop yield loss. At larger geographical scales, long-term meteorological data, such as precipitation, are taken into account to optimize crop area, crop choice and irrigation investment levels: model results suggest significant gains if implemented in most regions subjected to too much or too little water.

2. Fertilizer Application

More precise and less frequent fertilizer applications can strongly reduce costs and pollution of soils and water bodies, without negative effects on yields. By using big data tools such as spatially and temporally explicit N-cycle models coupled with soil and weather data at large scales, farmers can make more precise decisions on the timing and level of fertilizer application that accounts for the significant temporal and spatial variability in crop nitrogen requirements in high-use input regions. Further, based on the identification of the specific nutrient mineralization patterns from the long-term monitoring of soil properties through big data tools such as near real-time mapping and monitoring systems, farmers can optimize the fertilizer amount, timing and method of application to different parts of a field. Such nutrient optimization can be coupled to precision technologies that redistribute the right nutrient types per precision area.

3. Pest Control

Big data tools can help optimize pesticide decisions in terms of timing, type, and volume of pesticides at farm and at larger scales to increase efficacy and decrease costs. Climate, weather, pest cycle, and crop data can be analyzed at the regional scale to forecast and timely report pest outbreaks, which can be used by farmers to monitor their own fields. At the farm scale, big data tools can help farmers identify specific field areas in need of pesticide treatment, potentially coupled with smart spraying devices. Such early timely pest alerts or targeted pesticide advises can be effectively sent via mobile phone apps to producers or farm managers. At a larger geographical scale, lagged precipitation data have been used to improve seasonal forecasts of yellow rust.

6.7.1. Water Management

How to manage water efficiently? Only with precision irrigation can we use water as sustainably as possible. For that, it should be applied in the right amount at the right moment. The question is how, namely: which system to use, what are the climatic needs of the crop, what is the economical way to schedule the water applications, etc. Those questions should be asked by many people, and one of them is the soil which tells us its actual condition (dry... wet, salty... not salty, etc.) and its needs.

In principle, we have two types of sensors: 1. Ground truth measurements, which can be used to calibrate the models and the algorithms using the sensor data, and illuminate the sensor data being used posteriorly; and 2. Some models: soil plant models to model a plant's need for water, computation models to compute the reference ET0, empirical, semi-empirical, or physically based models to compute a soil status in water at a given time.

Whatever the option that we choose, without a good calibration of the model or the algorithm at hand, we will not receive a confident answer. In that line, any model or sensor will need good calibration. Agronomy has already developed its own systems for computing plant needs and soil conditions. Satellite-based power and mass exchange models produce evapotranspiration data, soil wetness models generate soil wetness data, crop phenological models generate crop developing state or canopy cover data, crop yield models generate yield components data (theoretical viable yield quantity, yield constraints, harvest index, yield response to water stress, growth degree-day rainfall core project), which can be validated, calibrated, assimilated, managed, or fused by ground truth measurements to improve accuracy and precision.

6.7.2. Fertilizer Application

Fertilizers are a key component of crop production and contribute to the sustained yields achieved in modern agriculture. However, the usage of fertilizers has to be optimized since excessive fertilizer applications can cause yield loss with no economic return whatsoever. Furthermore, water pollution reservation problems from leaching has caused soil degeneration by making it less resistant to erosion and salinization, making it less viable due to acidification, and compromising water quality. Throughout the world, in land used for farming, over 90% of biological nitrogen has been allocated to supporting crops. The objective is to focus on four practical applications: determining the technological resources for optimum yield and the location of these resources; the relationship between where fertilizers should be applied to a field to maximize profit, and where the crops will respond to fertilizer inputs; and improving the representational capability and the speed of soft computing alternatives to simulate crop yield.

To optimize fertilizer input, an understanding of the field variability during the various stages of development is required. This has been addressed in precision agriculture in different ways, taking into consideration adaptive zones with similarity analysis of different sites or multiple sensors indispensable for specific measurement tools throughout the growth season. Fertilizer application modeling has usually relied on developing experience curves giving return on investment for the average yield goal or other criteria historical estimates based on factors of performance, such as soil texture and climate conditions.

6.7.3. Pest Control

Pest and disease protection in crops has in large part relied on synthetic chemicals. The advent of crop biotechnology, particularly genetically programmed crops resistant to diseases and pests, has provided more options to the farmer to control crop health. Nonetheless, a careful application of the tools available should provide the optimal return for society of pest and disease protection in crops. Recent advances in pest and disease models and mapping and monitoring technologies can serve to optimize the use of available resources for pest and disease protection in crops. Data monitoring equipment deployed in the field can provide crop stress indicators useful for pest inherent capacity assessment. These models have been coupled with pest and disease establishment risk models to provide the farmer with indicators and maps to make the best decision to prevent flood related production problems with specific stored commodities. Testing pest detection procedures has confirmed their application potential.

Climate change is increasing the pace at which plant pests and diseases are modifying their pathogenicity, utilizing chemicals in agriculture more wisely to limit the development of resistances, taking advantage of available field data in real time to apply detection tools with geographical information systems should help minimize losses from pests and diseases. Implementing nematode analysis in soils for crop rotation in precision agronomic management has demonstrated the increase of crop yields. Underground pest and disease risk modeling for potato near decay and other specific diseases has also demonstrated practical application potential with additional tools managing technology and precision field equipment.

6.8. Enhancing Farm Efficiency

Agriculture is a complex and highly demanding activity. Farmers have to take important decisions every day, from selecting the crop to deciding on the most suitable disease prevention strategy. In many occasions, these choices are taken without the availability of supporting information. For instance, the crop cycle, from sowing to harvest, can last months or even years, making it difficult to adapt the strategy to changes in customer demand. Forecasting the yield of the crops can help farmers to succeed in the choice of the harvest period, product quality and storage time, among others. This information helps in the implementation of precise production strategies that minimize crop waste, in what is termed as “precision agriculture”.

Precise supply chain management for agricultural products requires forecasting methodologies different from those applied in other industries, as in the agriculture case, products are not processed by adding value but by removing it. Agricultural supply chains have completely different characteristics than those of processed food. First of all, the products are bulk and as such the product quality – for instance the size and color of the potato – is an important input in the supply chain configuration. The large amount of product transformed requires transport operations that are planned with a longer lead time than those required in the case of processed food. The perishability of food makes it imperative to efficiently manage the logistics operations to minimize the waste in the journey from farm to fork. This is especially true for fresh products during the summer when high temperatures increase the risk for spoilage. The availability of consumer-related data coming from retailing or wholesale makes it easier to plan the sales of short life core products with a high level of precision.

6.8.1. Yield Prediction

Studies show an increased dependence on Earth's lifespan, thus demanding the maximization of agricultural production. However, to understand how to predict crop yield beforehand and ensure a profitable harvest, farmers have always required external elements that affect crop growth assessment. Making yield predictions is extremely valuable, as they warn producers about price changes and possible shortages, so that they can modify their plans and avoid economic crisis for themselves as well as society. With an increasing global population and dwindling crop resources, increasing crop productivity is a major task for scientists and researchers. Higher crop yield predictions are particularly challenging, because the actual yield of the crop depends upon many interacting factors, such as weather conditions, genetic background of the crop, pest and disease infestation, weed competition, site characteristics, water availability, crop growth management, market price, and political support for the agricultural economy, etc. Even the best computers cannot evaluate all of the conditions that predict crop yield; however, they can be trained to make accurate decisions concerning crop yield prediction using big data analysis technologies. Utilizing enormous crop production data, the prediction of crop yield has made huge advancements in the past decades. These data can be utilized for both deterministic modeling and data-driven forecasting. In the data acquisition era, terrestrial data are made rooted with sensors that are deployed on-field. These remote data can be obtained using satellite imagery and unmanned aerial vehicles. By retrieving spatial and temporal data from diverse sources with different resolutions, researchers attempted to combine them and improve crop yield prediction.

6.8.2. Supply Chain Optimization

One of key fields, in which applications of Big Data analytics, is supply chain modeling. The SC modeling starts with the farm supply modeling. These models correlate inputs and yields, and can be applied for demand and resource management stocks. For the SC supply, optimization models depend on input prices, yield and quality prices, and transport costs. On the first stage, they can estimate required transport capacities, turnover times, demand and quality services. On the second stage, they can help to optimize the inputs and the transport. On the third stage, they simulate yield responses on transport chain flows. On the final stage, they help to forecast future dynamics of the inputs, products, and constraints. Agribusiness SC optimization models recognize its structural composition, which is the key problem. The optimal SC design depends on the scenario.

In the multi-model SC design approach, basic SC with the common principles can be unified, while specific channels can be considered separately. Adapted multi-item models allow planning at the operational level supply of specific products within the centralized multi-product decisions. The own models design is essential for transport-distribution processes science. At the strategic level, the SC structures and functions are hardly studied, including integration and surplus allocation. On the tactical level, research concerns about the redundancy calculation and locations optimization. Most urgent is the SC modeling in the decision support for export and export-supporting tasks. There is a lack of multi-agent SC models with horizontal arrangement, including ports. Recognition of innovative partnership links is a priority, what serves for reduction of transaction and economic surplus distribution costs. The priority functions are input exchange, financial service, and lost illness compensation, ecologically friendly products.

6.9. Conclusion

Our study provides an overview of existing work in the area of agricultural data analytics, and highlights several areas for future work that fall under two complementary themes: the need for better services and tools for supporting analytics in agriculture, and further exploring new and existing ideas for analyzing agricultural data. These themes relate closely to the debate in the applied machine learning community, and to the general cry for deeper, or more mechanistic modeling, versus the opportunistic story of using black box tools from machine learning. In this chapter, we argue that both empirical testing and mechanistic modeling are important and relevant for agriculture, and then use data from our local region and consider specific services and tools that could address the large potential for significant local to global impact of success of applied agricultural data analytics at scale.

Our intent in offering this overview is to inspire and help coordinate efforts in the applied data analytics community towards the important demand for impact through agriculture analytics and decision support. The potential to greatly enhance the impact of prediction modeling and its delivery through context-appropriate tools on farmer productivity, sustainability and resulting climate both locally and globally is clear, and significant efforts towards this goal are underway across the globe and with a regional focus. Should this work result in the broad engagement of farmers in the use of relevant techniques for driving on-farm decisions related to input use and the interventions to mitigate unproductive fluctuations in output yield, it would result in mutual benefit to food security, climate change and technological impacts.

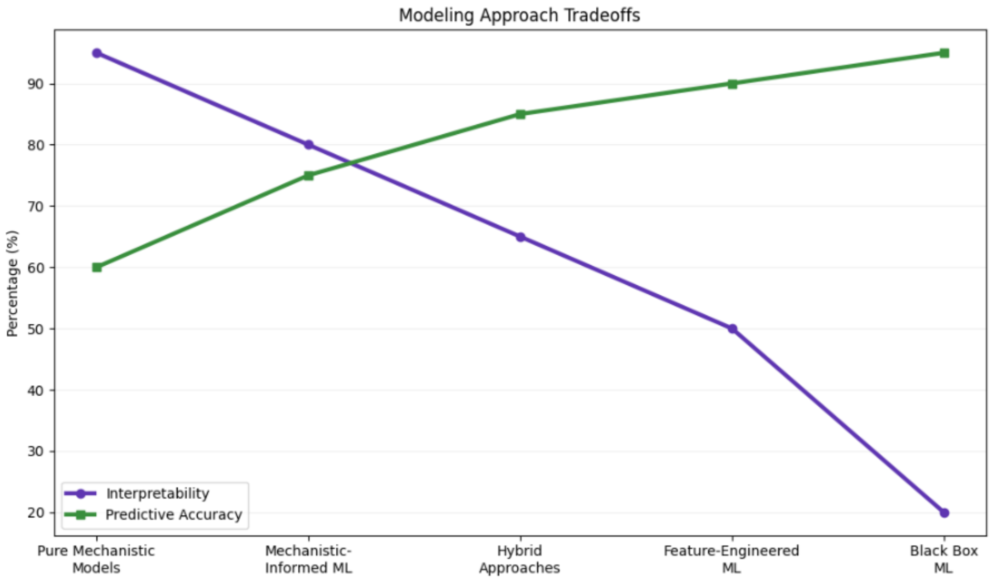

Fig 6 . 3 : Modeling Approach Tradeoffs

6.9.1. Key Takeaways and Future Directions in Agricultural Data Analytics

With the rapid development of information technology, data from agricultural production has attracted increasing attention from researchers and entrepreneurs, becoming one of the driving forces for the transformation and upgrading of agriculture. Based on the motivation of big data in agriculture, we reviewed the sources and types of agricultural data, the current status of agricultural data acquisition and processing, discovered new technical demands, and summarized the key technologies to support agricultural data processing. Then, a multilayer framework was proposed to provide possible support for higher agricultural data processing and driving processes of data-driven agriculture. Eventually, we presented a vision of the future sustainable agricultural data-driven ecosystem. Agricultural activities produce data that can be used as important supporting evidence for data-driven agriculture but unfortunately are not being fully utilized. In this sense, unexploited agricultural data from on-site services of agricultural processes can be different from previously published big data studies.

Moreover, the collection and use of agricultural data can be achieved based on a constructed agricultural data-driven ecosystem. This ecosystem requires various participants to share responsibilities in the acquisition, collaboration, communication, and decision-making support processes.

Driven by the concepts of intelligent agriculture, smart agriculture, and precision agriculture, the construction and application of the agricultural data-driven ecosystem must serve the driving of profitability for agricultural entities, whether small- and medium-scale family farms or large-scale plantation structures. With pasture production as an example, a closed agricultural data processing loop can be proposed. The core driving force for closed-loop operations is the agricultural parties on the production side, who can make feedback after paying for the personalized services or products designed according to their own requirements based on data analysis. Multiple-party and multi-round interactions can drive sustainable profits in the agricultural data-driven ecosystem.

References

Liakos K.G., Busato P., Moshou D., Pearson S., Bochtis D. (2018). Machine Learning in Agriculture: A Review. Sensors, 18(8), 2674. https://doi.org/10.3390/s18082674

Klerkx L., Jakku E., Labarthe P. (2019). A Review of Social Science on Digital Agriculture, Smart Farming and Agricultural Innovation Systems. Agricultural Systems, 184, 102763. https://doi.org/10.1016/j.agsy.2020.102763

Verdouw C., Wolfert S., Tekinerdogan B. (2021). Data-Driven Agriculture: New Opportunities for Smart Farming. Computers and Electronics in Agriculture, 181, 105964. https://doi.org/10.1016/j.compag.2020.105964

Kamilaris A., Prenafeta-Boldú F.X. (2018). Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Zhang Y., Wang G., Wang J. (2021). Applications of AI and Big Data in Smart Agriculture: A Review. IEEE Access, 9, 75999–76020. https://doi.org/10.1109/ACCESS.2021.3083991