Big Data Achieve's Sustainable Farming

The world's population is increasing day by day and most of this increase is occurring in developing countries. It is expected that the future income levels of the world's population will keep rising of what they are now; and to feed this larger and better-off population, food production must also massively increase. Therefore, agricultural research and development should be one of the prime focus in developing countries, with investments coming from both public and private sectors, ensuring smallholder farmers to get timely advice on their queries and access to new agriculture techniques and technologies. This article points to a few of the key obstacles for developing counties in adopting digital agriculture technologies; and argues that the application of big data to the agriculture sector in developing countries is beneficial to achieve sustainable and increased food production.

1. Introduction

Agriculture in developing countries contributes a significant portion to national GDP and provides livelihood to a large population. Despite agriculture-related technology advancements, the sector still faces structural weaknesses and challenges that include, but not limited to, partial availability of quality datasets, pests, climate vulnerability, and inadequate farming practices. The lack of adequate support for farmers to adopt good agricultural practices is yet another factor that obstructs the productivity. Farmers need up-to-date advice on crops' diseases, crop patterns, and adequate prevention actions to face developing circumstances. Currently, in developing countries, farmers' access to such information is extremely limited due to the agriculture support system being either not available, inconsistent, or unreliable/not timely. In this short article, we highlight the challenges and opportunities related to the adoption of big data applications to achieve sustainable farming.

The following sections are organised follows: a few of the related challenges faced by agricultural farmers are presented in Section 2 and various types of agriculture datasets are elaborated in Section 3. When deciding about agricultural data storage and processing, there are two most widely available options to select from, i.e., SQL and/or NoSQL. In Section 4, the factors based on which an informed decision about the selection of agricultural data storage and processing tools can be made are discussed. The discussion is then moved towards methods of collecting farmer's complaints and their resolution in Section 5. A few of core agricultural data analytics and decision support requirements are discussed in Section 6, and sector readiness and affordability are briefed in Section 7. Finally, Section 8 concludes the discussion and highlights a few of the challenges and opportunities of Big data application to achieve sustainable farming.

2. Challenges for agricultural farmers

Although farmers are involved in exhausting work to cultivate crops, they face numerous difficulties and do not produce up to the full potential. This usually due to the unavailability of the latest information and/or weaknesses in their agricultural practices. In order to build an agricultural information system for farmers, it is important to understand their current data sources and then identifying weaknesses, limitations, and challenges involved in efficient system development. These weaknesses may include not only managerial weaknesses and inadequate farming practices, but also technological gaps. The managerial issues and inadequate farming practices are mainly due to lack of updated knowledge, which results in unsuitable (or wrong) crop cultivation at the wrong time without considering the nature of the soil, or weather conditions, or market demand of crops in a given season. Another critical issue for farmers is the high risks of epidemiology, plant diseases, and pests due to the absence of timely actions that can be taken for their prevention.

These issues exist because farmers typically follow the practices of their ancestors or may ask friends that are inadequate/outdated or get information from television that telecasts news and advertisement irrespective of the farmer's regional needs or scenarios. A large number of agricultural software are developed to-date in order to educate farmers with technological information. However, most of them provide static information about farming or require a large number

of searching steps to get accurate information. These systems do provide information on seeds, crop diseases, weather advisory, optimal harvesting times, etc.; however, farmers are usually unable to get benefit from them due to unrelated data, diverse information or their complexity.

In order to guide the farmers and prevent them from significant losses, many countries have established agri-centres at major rural hubs. Farmers take guidance from the domain experts sitting at agri-centres. Most of the existing farmer guidance systems work manually, whereby the experts sitting in the agri-centres advise the farmers in case of any query/complaint through field visits. Moreover, the arable lands are stretched throughout the country, but the number of agri-centres is limited, with a limited number of domain experts available to perform field visit surveys for farmer queries and complaints, to make timely analysis, and to then provide useful recommendations. Another issue is because of the lack of historical data, availability of digital platforms, and pieces of training for experts that adversely affect the efficacy of their recommendations.

3. Agriculture data types, collection, and storage

The data that is related to the agriculture sector can be processes related, crops metadata, machine logs, human-sourced, historical data, etc. The process-related data is usually generated by data records of seeds, soil, fertilizer, etc. Crops metadata is usually about crop categories, types, and subtypes. Both the process generated data along with the agriculture metadata is mostly structured or semi-structured. It usually includes data related to entities and transactions. Machine logs are data collected using smart devices in real-time by the agriculture sector; for example, using soil sensors that measure temperature, humidity, sunlight, crop health monitoring systems, automatic yield recording systems, irrigation facility monitoring systems, etc. The technologies like the Internet of Things (IoT) allow real-time data to be accessed by farmers. Human sourced data involves social media feeds and trends on food production and consumption. These machine-generated data and human-sourced data are usually semi-structured or unstructured. The historical data could be related to historical agricultural practices and/or the data generated from the farmer's complaints/support that follows a conventional complaint submission approach where farmers submit their complaints and then advice or solution is suggested by the respective regional agricultural support centres. This historical data could be from different sources and a mix of all structured, semi-structured, and unstructured data. Since all of the above-mentioned agriculture-related data is large, scalable, heterogeneous, and doesn't conform to a single scheme, it poses various challenges in the collection, processing, storage, and in making use of this data. Therefore, agriculture sector embarking on big data projects need to choose which data storage to be used, and often that decision swings between SQL and NoSQL. SQL has an impressive history with a huge customer base, but NoSQL is making many notable expansions too. In this regard, the next section highlights a few of the key factors to consider when selecting SQL or NoSQL for agriculture data storage and processing.

4. Factors to consider when selecting SQL or NoSQL for data storage

In general, there are numerous agriculture-related dataset types with diverse processing requirements. This data can be broadly related to two aspects, i.e., transactional data for recording day-to-day operations and historical data for running data analyses queries. The transactional data may include inventory management such as carrying out stock updates in real-time during harvesting period or recording suppliers, sales, or payments - which can be managed by a RDBMS. Agriculture data analyses could involve various types of datasets collected from a variety of sources; for example, the relationship between soil and crops, supply chain, pesticides, etc., which could be useful to make production decisions. Moreover, historical data can be combined with these data sets to forecast crop yield, demand, and to prevent agriculture loss. Processing and analysing this massive amount of structured, semi-structured and unstructured data using a relational database system can be difficult. To deal with such large volumes of data that have little to no structure, NoSQL databases are more suitable[1]. This is mainly because NoSQL complies with BASE (basically available, soft state and eventually consistent), and they do not limit the types of data that you can store together. Using NoSQL can give flexibility, scalability, and time-efficient data processing. This can be achieved by using the Map-Reduce software framework for processing vast amounts of data using a large number of computers (nodes), collectively referred to as a cluster. In summary, the use case of NoSQL is different from the standard SQL, which offers transactions in a tabulated format. NoSQL databases have proved to be flexible and can provide a better ad-hoc query feature for large datasets.

5. Farmers complaints and resolution

Over the last few years, most of the agricultural-related research and development was toward various aspects of precision agriculture, irrigation management and optimal farming by developing decision support systems (DSS) and data analytic tools. These systems are extremely useful in providing structured information in a step by step manner. However,

there is still lack support for rural farmers in terms of addressing their regional queries, questions and complaints. There is a dire need to have near real-time automated complaint management systems that can provide an interactive way for farmers to access information and submit complaints based on their unique preferences. Over the years, several systems have been developed to resolve complaint issues faced by farmers. Social networks and interactive voice response (IVR) systems are also developed for farmers where they interact and help each other[2]. Governments establish agri-centres at major rural hubs where domain experts advise and resolve farmer's queries through telephone calls. However, it is not practical to manually handle all the calls and to keep up with the high user demand, especially in keeping track of the queries that farmers raised earlier or have another issue with the same previous context. Also, it is difficult for domain experts to provide accurate response against farmer's queries and complaints only through conventional phone calls as full information regarding the plants and underlying issues may not have been appropriately communicated.

In a typical scenario, farmers manually submit complaints to their respective 'agricultural associations' on a structured paper-based form. These complaints are then sent to one of the agricultural centres distributed over the country to offer support to the farmers. Several agricultural experts working at these centres subsequently process farmers' inquiries. A recommendation is then provided, or in some cases, a 'no known solution' is delivered – usually via phone calls. Even with a swift round of consultancy provided by the system, response from experts gets significantly delayed, mainly due to a large number of received queries (tens of thousands). Consequently, farmers mostly get an answer when it is too late for them to act. Equally important, the support provided by experts deals only with farmers' instant complaints, lacking near future perspective on developing circumstances, and thus advice for the upcoming scenarios. Designing and developing an efficient query/complaint management system for sustainable farming is still an open problem and pressing issue. In an effort to improve this, Mohit Jain et al.[3] considered Google Translator and IBM Watson Speech-based conversational system to propose a conversational agent for resolving farmer queries. Further improvements to such systems can be made by utilising multimedia technology to enhance the usability and acceptability aspects for farmers with limited literacy while keeping it highly scalable, available around-the-clock and have manageable overheads.

6. Agriculture data analytics and decision support

The latest advancements in the IoT technology, data collection tools and agriculture robotics are being adopted by the agricultural sector for smart farming. These technological gains are steering the traditional agricultural practices in an era of digital agriculture that enhances productivity. Consequently, the large amounts of data being collected are expected to have a positive impact not only on smart farming but also on unprecedented decision-making capabilities to the farmers and government. The agriculture big data analytics allows patterns and relationships to be found in the data, enabling farmers to make predictions about crops and/or agricultural land[4]. For example (a) farmers can decide what and when to grow by analysing the relationships between crops, land and weather conditions; and (b) if farmers can know which crops are likely to be in demand, they will be able to decide in advance what to grow and give themselves a huge competitive advantage, etc.

Agriculture data analyses can broadly be divided into four levels of decision making i.e. descriptive, predictive, prescriptive and proactive. Here, descriptive decision making includes decision like efficient production to know things like when and where to plant. Such analyses may include data related to soil, weather, supply chain, pesticides, etc. The predictive decision making mainly involves historical data combined with descriptive data to generate future forecasts such as production levels and food uncertainty. Such data can be used to anticipate crop production, demand for seeds, fertilizers, labour, etc. The prescriptive data analysis is used for making decisions regarding intervention, e.g., to identify steps that need to be taken at an early stage to prevent agriculture loss. Finally, the proactive analysis in the agriculture sector is usually related to observing the development of crops over time to understand relations between location, seeds, fertilizer, weather and crop production. However, in order to perform useful data analyses, farmers may need to perform data pre-processing to improve data quality. The amount of effort required in improving data quality is variable, depending on various factors, e.g., sources of data, data collection methods, relationships or associations among datasets etc.

7. Sector readiness and affordability

Over the past few years, the adoption of the latest technologies and use of Internet of Things (IoT) in agricultural research and development has rapidly increased. All the agriculture-related data being collected through these developments, and processed through big data technologies, have huge protential to bring the next revolution in agriculture. One such example is Precision Farming or Smart Farming, which is becoming a key inclination in the developed and/or industrialised countries. In many of these countries, the agricultural data is being analysed with detailed records on

historical weather conditions, landscape, and crop productivity combined with real-time IoT sensors. Here, the concept of Smart Farming or Precision Farming is about managing the agriculture farms with the latest technologies and infrastructures, including big data, cloud computing and IoT – for managing, automating and analysing farm operations[5].

Despite all rapid advancements, and especially the increasing use of information and communication technology for agriculture, a lot is remaining to be done to ensure smart farming is successfully achieved. In this regard, developing countries are still lagging far behind due to various reasons, including the digital divide. The first challenge faced by these countries towards the employment of smart farming in rural areas is the absence of internet/telecom connectivity. Moreover, the availability of historical agricultural data is practically the most critical issue in developing countries. There is an absence of national and regional agricultural data collection centres; and if they collect, those data sets are not shared with the farmers due to the security and privacy concerns.

In a nutshell, the key obstacles for developing counties in adopting latest decision support and big data technologies for agriculture can be pointed to unavailability of digital skills, high costs of technology and infrastructure for farmers, poor national telecommunication infrastructures, no access to national or regional historical agriculture data, concerns on data privacy and ownership, undeveloped data sharing policies, etc. Majority of farmers (and especially the smallholders) residing in developing countries are faced with several (if not all) of these challenges. Therefore, for such countries, the trend towards Smart Farming needs to begin with small agricultural development projects. For example, at the time of writing this article, agriculture in Egypt was absorbing over 30% of the workforce and providing livelihood to over 50% rural population, but agriculture contributed around 11.05 per cent to the GDP of Egypt in 2019[6]. One of the main reasons behind this is that each year, a large portion of crops are wasted due to pests, diseases and inadequate farming practices. It is believed that agricultural projects investigating systems for timely farmers' complaint resolution, and access to information and expert advice are the key initial steps to achieve sustainable and quality agriculture production in developing countries like Egypt. One such effort is the AgroSupportAnalytics (a Cloud-based Complaints Management and Decision Support System for Sustainable Farming) project[7]. The AgroSupportAnalytics project aims to resolve the problem of support and advice for farmers in the current manual system by developing an automated complaint management and decision support system. It is being based on the application of knowledge discovery and analytics on agricultural data and farmers' complaints, deployed on a Cloud platform. The proposed automated system will be used to provide adequate and timely advice for farmers upon their enquires / complaints, and also to foresee near future development of circumstances by the experts[7]. Consequently, enabling agricultural experts to broadcast early warning signals of threats, mainly pests and disease, and the needed prevention actions to be undertaken by farmers. The system is currently focusing on improving support for farmers on the farming of Wheat, Rice and Cotton – major field crops in Egypt.

8. Conclusions: challenges and opportunities

Agriculture in developing countries contributes a big portion to national GDP, but there is a lack of effective support for farmers to adopt suitable agricultural practices through technology advancements. Farmers usually require timely advice and suggestions on crop patterns, diseases and prevention actions to tackle emerging situations. However, farmers' access to such information, especially in developing countries, is limited due to the agriculture support system being either not available or not fully automated.

The data that is related to the agriculture sector can be processes related, crops metadata, machine logs, human-sourced, historical data, etc. Processing and analysing these massive amounts of data is challenging as they are usually collected from different sources and can be structured, semi-structured or unstructured. In this context, a critical decision in developing big data agricultural systems is based on the selection of data storage and the decision swings between SQL and NoSQL, depending on the nature and modalities of data involved in the project.

Smart farming involves managing the agriculture farms with the latest technologies and infrastructures, including big data, cloud computing and the internet of things (IoT). The large amounts of data being collected in the agriculture sector are expected to have an impact not only on smart farming but will also improve the decision-making capabilities of the farmers and government.

The future of agriculture undoubtedly seems to lie in embarking on big data technologies and smart farming because of their flexibility, scalability and innovation. The adoption of big data tools and processes is expensive and requires a lot of technical skills, so bigger businesses are most likely to benefit from them quickly. But this is unlikely to happen quickly for most small farmers in developing countries who are poor and also not quite ready for a radical change. Moreover, the unavailability of digital skills, poor national telecom infrastructures, no availability or access to national or regional historical agriculture data, etc. are few obstacles in developing counties to adopt the latest technologies. However, if

efforts are made at the government level and by the national agricultural centres for the development of small to medium size projects; such as the AgroSupportAnalytics project[7], in order to improve access to agriculture data, development of online complaint management and big data decision support system, etc, then small farmers in developing countries can equally benefit from the emerging technologies.

References

1. Papajorgji, Petraq; Pinet, Francois; Innovations and Trends in Environmental and Agricultural Informatics. *Advances in Environmental Engineering and Green Technologies* **2018**, *2018*, 38-57, 10.4018/978-1-5225-5978-8.

2. Waleed Riaz; Haris Durrani; Suleman Shahid; Agha Ali Raza; ICT Intervention for Agriculture Development. *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* **2017**, *17*, 1-5, 10.1145/3136560.3136598.

3. Mohit Jain; Pratyush Kumar; Ishita Bhansali; Q. Vera Liao; Khai Truong; Shwetak Patel; FarmChat. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2018**, *2*, 1-22, 10.1145/3287048.

4. Hassina Ait Issad; Rachida Aoudjit; Joel J.P.C. Rodrigues; A comprehensive review of Data Mining techniques in smart agriculture. *Engineering in Agriculture, Environment and Food* **2019**, *12*, 511-525, 10.1016/j.eaef.2019.11.003.

5. Sukhpal Singh; Inderveer Chana; Rajkumar Buyya; Agri-Info: Cloud Based Autonomic System for Delivering Agriculture as a Service. *Internet of Things* **2020**, *9*, 100131, 10.1016/j.iot.2019.100131.

6. Egypt: Distribution of gross domestic product (GDP) across economic sectors from 2009 to 2019 . Statista. Retrieved 2021-2-10

7. A Cloud-based Complaints Management and Decision Support System for Sustainable Farming . AgroSupportAnalytics. Retrieved 2021-2-10