Application of Data Warehouse and Big Data Technology in Agriculture in India

N P Singh
Management Development Institute
Mehrauli Road, Sukhrali,
Gurgaon-12207, India
E-Mail: Knpsingh@mdi.ac.in

DOI: 10.1481/icasVII.2016.f29

ABSTRACT

In the recent years, it is observed that the scientist, planners, executives across the globe are using data collected from traditional record keeping by government agencies, data collected using sensors and satellite imagery technologies and combining it with predictive weather modelling. It is being done to help farmers to make better decisions with respect to optimal time of sowing/planting of the crops, optima time for application of pesticides, insecticides, and fertilizers starting with sowing, and time for harvesting crops in general and especially in a given region of the country at a given point in time. The technologies such as data warehouse/ big data are also being implemented to improve the quality of data since data provided by different arms of the government such as organizations of central government and state government varies a lot on many parameters and create problem in making right decisions. This is being done as part of digitization India program of central government. Keeping in view, the importance of quality and accuracy of data for planners, researchers, political leadership and farmers in particular, this paper presents various aspects of the data warehouse/ big data technology application in agriculture in the context of India. These aspects are (i) Identification of major domain of implementation of data warehouse/ big data technology in agriculture in India, (ii) identification of present status of application of data warehouse /big data technology and harvesting the benefits of the technology in India, (iii) identification and quantification of the benefits of the data warehouse/ big data technology application in increasing agriculture production or reducing losses to farmers (iv) identification of most popular data warehouse / big data technologies to be implemented in India (v) identification of targeted components of crop life cycle for data warehouse/ big data technology in agriculture in India, and (vi) impact of usage of technology and role of technologies such as data warehouse /big data in eradicating hunger or poverty from earth by way of helping farmers in increasing production and in maintaining ecosystem of the crops. The analysis being carried out by keeping in view irrigated and rain fed areas under cultivation in India and also changing climatic conditions which are impacting agriculture production in India.

Introduction

India's total area is 3.287 million km². The area under agriculture was 60.6% during 2013-14. Agriculture sector (Agriculture proper & Live stock, Forestry & logging, Fishing and related activities) accounted for 17.8%[1] of the GDP (gross domestic product) in 2013-14. It was more that 51% during 1950-51. The total production of agriculture sector was US$ 366.92 billion. Indian share in world agriculture output is 7.68%. About 50% of the workforce is employed in agriculture in India. Though the contribution of agriculture to GDP is steadily declining but still agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India. As per FAO 2010 reports, India was the world's largest producer of many fresh fruits and vegetables, milk, major spices, select fibrous crops such as jute, staples such as millets and castor oil seed. India is the second largest producer of wheat and rice.

Food security of India depends mainly on production of cereals crops. However, it needs to increase production of fruits, vegetables, pulses, and milk to meet the demands of growing population with rising income. India's priority areas in the context of agriculture are (i) higher productivity with optimal usage of water and other resources, (ii) poverty alleviation and management of natural resources, (iii) sustaining the cultivated land for future productivity and environment, (iv) reducing the waste of food grains and other agriculture produce, and (v) keeping the balance of prices of agriculture produces between producer and consumers, and many more.

The key to success depend on availability of quality data. The quality data can be captured / created, stored in times series manner, and analysed with technologies such as data warehouse, big data, data analytics, and reporting tools. To mention here as an example, Big data in agriculture refers to the Electronic Farm Records (EFR) which includes soil temperatures maps and data, precipitation maps and data, electrical conductivity maps and data, moisture content data, air permeability maps and data, nutrient contents and pH level data, past cultivation records, past losses, insurance and yield related information. It also includes social media data including tweets, blogs, new feeds, agriculture research reports and research articles in agriculture and other related journal. The technology will also help in mining the big data and discover the associations, understand patterns and past & future trends to improve the agricultural systems, accurate estimation of large number of input and output parameters, increase crop productivity in relation to changing input parameters, and reduction in input costs by accurate diagnoses and analysis of large numbers of factors effecting it.

Keeping in view the importance of quality data as suggested by Addison (2015), the present research paper is an attempt to (i) identify major domain of implementation of data warehouse/ big data technology in agriculture in India, (ii) identify present status of application of data warehouse /big data technology and harvesting the benefits of the technology in India, (iii) identification and quantification of the benefits of the data warehouse/ big data technology application in increasing agriculture production or reducing losses to farmers (iv) identification of most popular data warehouse / big data technologies to be implemented in India (v) identification of targeted components of crop life cycle for data warehouse/ big data technology in agriculture in India, and (vi) impact of usage of technology such as data warehouse /big data in eradicating hunger or poverty from earth by way of helping farmers in increasing production and in maintaining ecosystem of the crops.

[1] http://data.worldbank.org/indicator/NV.AGR.TOTL.ZS

The analysis being carried out by keeping in view irrigated and rain fed areas under cultivation in India and also changing climatic conditions which are impacting agriculture production in India.

Research Methodology

The present study can be termed as exploratory cum descriptive in nature. Research article is based on secondary as well as primary data. Secondary data is about application of data warehouse and big data technologies in agriculture and primary data collected from 20 villages of Aligarh District of State of Uttar Pradesh through group discussion method from farmers. This data was mainly about usage of machinery by the farmers, challenges of supply chain of agriculture produce, government initiatives, and their awareness about data collection by different agencies with respect to soil fertility, applications of fertilizers & pesticides and other inputs, existing practices adopted by farmers etc. The data is analyzed mainly in the form of frequency distribution.

Result & Discussion

Identification major segment of data warehouse/big data technologies in agriculture in India (1).

Big data / data warehouses technologies along with mobile technologies can be used in sourcing large set of data of agriculture sector activities, processing of this data, and generating valuable information for farmers, consumers, and government agencies from large set data. To be specific, Big data farming (or precision farming) is going to help farmers by offering to farmers (i) highly accurate weather forecasts and real time field data to optimize the resources & to reduce losses to farmers specifically in rainfed regions, (ii) real-time optimization of farming machinery, irrigation schedules, and other inputs, (iii) cloud-hosted cost effective information resources for farmers, (iv) automated irrigation and other farm practices recommendations, (v) monitoring the attack of insects, pest & diseases on crop with changing climatic conditions, (vi) monitoring of prices of agriculture produce before sowing and after harvesting, (vii) management of inventories of public distribution and its large & complex supply chains, (viii) monitoring of temperature, humidity, oxygen levels, and other needed parameters of food storage containers and shipment containers to maintain quality of agriculture produce before it is consumed, (ix) collection and analysis of post harvest field data for suggesting the activities with respect to maintaining the fertility of soil and making agriculture environmental friendly.

Identification of present status of application of data warehouse /big data technology and harvesting the benefits of the technology in India (2).

Four set of technologies are implemented in an organization to harvest the real benefit of information technologies. These are (i) technologies to source data/capture data/automate business processes, (ii) technologies to transport data from one place to another, i.e., from source to destination (Sylvester (2013)) (iii) technologies to integrate and store data, and (iv) technologies to analyse and report the data / information to consumer of information. Due to large size of the country as well number of organizations, implementation of all four set of technologies is not yet completed in India. In addition, implementation of technologies is easy in government owned organization but difficult in a segment which is not so organized. For example, meteorological department is collecting, analysing, and reporting information to users about meteorological parameters in general but not in relation to the cultivation of a particular crop in a particular region and its impact. Secondly, basic data of 105, 64,74, 292 citizens (80+% of population) is captured as part of Unique Identification Authority of India (UIDAI) project using data warehouse technology (https://portal.**uidai**.gov.in/). Thirdly, major procurement and distribution agencies such as Food Corporation of India and others are computerised for monitoring the inventory of food stocks on day to day basis (http://fci.gov.in/). There is a need to integrate various data bases in order to derive values to farmers out of this data.

Identification and quantification of the benefits of the data warehouse/ big data technology application in increasing agriculture production or reducing losses to farmers (3)

Many studies are conducted for identification and quantification of benefits of data warehouse / big data technologies. Madgavkar and Krishnamurthy (2014) estimated that technology based applications in agriculture in India will have US$ 45-80 billion impact annually as given in table 1. These estimates are not comprehensive.

Table 1: Impact of Technology based application on Indian agriculture (US$ Billion)

Sized Applications	Economic impact	Potential reach	Potential productivity or value gains
Hybrid and GM crops	1-4	10% of total of 92 million tones of farm produce under GM crops	5-10% productivity improvement
Precision Farming	8-30	20% of total arable land under precision agriculture	15-60% yield improvement 2ill help 22 million farmers.
Real-time market information	10-15	90 million farmers (60% of total) using real-time market information	3% productivity increase, 2.5% increase in price realization, input cost reduced by 3%.
Reduced leakage and Waste	27-32	$19 billion leakage in (PDS), $28 billion of non-PDS food waste	Up to 90% reduction in PDS leakage. 50% lower wastage in distribution of other produce.

Yield is a function of genetical characteristics of crops, environmental conditions & farm practices. The input variables are not independent. One need to develop predictive/ prescriptive models to optimize yield for a specific environment by optimizing farm practices for genetically different crops. This is possible only by conducted experiments, collecting and analysing data of these experiments. This will need in turn data warehouse/ big data technologies to store and analyse large set of data. It will result into many useful finding to boost agriculture at low cost. To mention, farmers in India usually burn residues of the rice crop after harvest in winter and of wheat in March-April. However, it is observed in adopted villages under Consultative Group for International Agricultural Research (CGIAR) project on Climate Change, Agriculture and Food Security (CCAFS) that zero tillage along with residue management and diversification of crops reduce the fertilizer requirement by a fifth after three years. Experimental data revealed that a tonne of rice and wheat residues, about 40% of which is carbon, contain 5-8 kg of nitrogen, 1-2 kg of phosphorus and 11-13 kg of potassium.
Further empirical studies confirmed an increase in yield by 10-15% with zero tillage and line sowing of wheat. Zero tillage reduces diesel cost by 80-85% (Seetharaman (2016)). This is not propagated in other part of the country due to non-availability of data in electronic form.

Identification of most popular data warehouse / big data technologies to be implemented in India (4)

Big data started with precision farming. Further, it can store data on fertilizer consumption (from production in factories and imports if any), planting (data are collated from different sources), crop protection (data are collated from different sources), harvesting and yield of crops (large number of crop cutting experiments & imaginary data). In addition data of 3rd parties on weather, data from satellite / aerial imagery (large amount of granular data is collected), data of soil fertility (not much data is collected), topographic data (large amount data is available with many companies), research & development data (data from academics & industry), land records and fertility data (data are collated from different sources), and data of machinery and equipments of manufacturers, data of commodity markets (local, regional and global), weather data can be integrated with big data technologies. Scope of these technologies is unlimited but major initiatives are needed (i) in storing data of all farm produces, (ii) data of live stock & fisheries etc, (iii) data of complete supply chain of farm produce, (iii) marketing and procurement, distribution policies and systems. It will help in reducing losses. Once it is achieved in a comprehensive way more initiatives should be taken to input side of supply chain.

To achieve these objectives government scientific organizations have taken initiatives in identifying utility and segments of agriculture sector for implementing these technologies. To mention, (i) Council of Scientific & Industrial Research (CSIR) has approached scientific community to work on research projects to prepare roadmap for big data technologies in India (http://www.dst.gov.in/big-data-initiative-1), (ii) National Agricultural Bioinformatics Grid (NABG) project at Indian Agricultural Statistics Research Institute (IASRI) launched in 2010, (iii) Advanced Supercomputing Hub for OMICS Knowledge in Agriculture – ASHOKA established during 2014. Peisker and Dalai (2015) suggested a frame work for agriculture sector once connectivity with households, villages, elected village level institutions, government departments is established. They reported that present sources of data are (i) kisan SMS portalsystem, (ii) Community information Center, (iii) AGMARKNET (http://agmarknet.dac.gov.in), (iv) e-choupal, (v) agriwatch.com. These sources may be further integrated and augmented. Channe et al (2015) proposed an approach based on five key technologies: Internet of Things, Sensors, Cloud Computing, Mobile Computing and Big-Data Analysis.

Precision Farming Development Centres (PFDCs) been established in India to promote "Precision Farming & Plasticulture Applications for high-tech horticulture" and located in State Agricultural Universities (SAUs); ICAR Institutes such as IARI, New Delhi; CIAE, Bhopal & CISH, Lucknow and IIT, Kharagpur. These centers have been operating as hub-centers of plasticulture and precision farming in respective states. National Committee on Plasticulture Applications in Horticulture (NCPAH) during the year 2008-09 established five new Precision Farming Development Centres located at Bhopal, Imphal, Leh, Ludhiana & Ranchi under the centrally sponsored scheme Micro Irrigation (http://www.ncpahindia.com/pfdc-mandate).

Identification of targeted components of crop life cycle for data warehouse/ big data technology in agriculture in India (5)

Data warehouse/ big data technology is needed during the complete life cycle of the crop but capturing data on all factors impacting yield is not possible with existing resources available with developed economies and specifically in India. With the existing data on rainfall, seed, farm practices, fertilizer, pesticides, insecticides production & imports, these technologies can target data on inputs and its analysis to extend full advantage of moisture content of soil as well as life in days of the crop. Second, life cycle to be targeted for data collection and analysis is post harvest period. Initial cycle will help in reducing cost of cultivation and post harvest will reduce losses of supply chain.

Impact of usage of technology such as data warehouse /big data in eradicating hunger or poverty (6)

This section presents problems of supply chain of agricultural produce and other related problems and the technologies used by farmers not necessarily data warehouse and big data technologies. These findings are based on primary data collected from Aligrah District (Northern part of India) mainly on three themes. The findings on these themes are listed in the following:

(i) Usage of farm machinery: Indian farmers have started using machinery at their fields. The popular ones are tractors, walking tractors, combine, laser levellers, and other farm farm equipment to open furrows in the ground, shredding, spraying and fertilizing the soil. In many cases machinery is replaced by services provided by farmers from State of Punjab. In the recent years due to increase in cost of diesel, almost all farmers have applied laser levellers and auto start switches for water pump sets. They are also using tractors, combine but not much is employed for data collection. Data collection is either manual or through imaginaries.

(ii) Supply chain of Agri-produce and challenges: Indian agriculture supply chain is spread in large geographical area. The agriculture produce are grown in specific geographical areas and sold by farmers either to government agencies, whole seller or retailers in other parts of the country. Road transport is major mechanism to transport agriculture produce for shorter distance and rail for longer distance. The agriculture markets are not yet developed in smaller towns, & district headquarters. Major movements are between large markets.

Based on analysis of data collected form 20 villages from Aligarh district (State of Uttar Pradesh) it was inferred that (i) every farmer brings its produce specifically grains, fruits & vegetables, milk etc to the centers listed in fig 1 (Khair, Atrauli, Khurja, Iglas, Manai, and Aligarh) with no advance information about demand-supply and price variations. Another group of consumers (farmers & non-farmers) come to these centers to buy the same produce from these towns at a high cost due to market taxes, and transportation costs etc at other time (ii) all storage facilities are located in these towns which in turn increases transportation costs and also loss of work at home, (iii) in many cases the farmers bring their produce to nearby state of Delhi which operates a larger markets and small traders buy the same produce from Delhi markets & do retailing in the small towns which are much nearer to producer farmers. It results in to higher transportation cost to traders & farmers, (iv) most cases return to the farmers are not linked to quality of the produce in many cases, as they cannot pack or certify their produce as organic and inorganic (v) farmers are guided by resources available, and advise of local traders or peers with respect to usage of fertilizers or pesticides rather than scientific methods, (vi) their knowledge about residual effect is very low, (vii) crop rotation is limited to wheat & paddy and in other segments potato & paddy, (viii) all farmers burn residues of wheat, sugarcane, paddy, maize with no interference from experts, (ix) farmers are not much aware of data collection activities by any agencies for the purpose of new research initiatives, and (ix) invest lots of resources in tillage which is not needed specifically in paddy.

Fig 1: Map of Aligarh District (Uttar Pradesh)

(iii) Making government initiatives more effective & beneficial to farmers.

Indian government protects the farmers by fixing minimum support price for major crops every year. Government agencies (Food Corporation of India, State corporations etc.) procure major food grains directly from farmers every year as part of its food security program to provide major food grain to the weaker sections of the societies though public distribution systems. Government has implemented information technology based systems for capturing data of the activities of these organizations but integration of data among these agencies is still needed. As mentioned, earlier all data of benefits of government schemes is linked to Aadhaar card (UIAI) and also to the bank account of farmers. It has reduced leakage of government monetary help to the farmers.

Concluding Remark:

It is beyond doubt that application of data warehouse / big data technologies will help all members of eco-systems. Quality data is vital for growth of agriculture in India. Government has taken many initiatives for capturing data and making use of it. The success is achieved in relation to collection and analysis of (i) weather data, (ii) forecasting area under different crops & yield using remote sensing application, data of crop cutting experiments, and data collected by state irrigation departments etc, (iii) data of procurement of food grains & vegetables, storage centers, public distribution systems etc, (iv) data generated academic/research institutions, (v) information about shelf life of the produce and many more segments of eco-system.

What is needed? Integration of all these data sets using data warehouse and big data technology for the purpose of analysis to make concept such as precision farming realty. Most of data used in research is generated from big agricultural universities and institutions farms where resources are available in developing new varieties etc but this is not the case with farmers. Experiment must incorporate some constraints to see the potential of new varieties or new practices. This will be possible in a country like India with private sector development in developing such technologies, public and private collaboration in implementation, and government support and investment as in case of some other sectors (World Economic Forum (2012)).

Reference

1. PTI (2016). Government to add 56,800 tones of storage capacity for onion, May 31. Retrieved on June 02, 2016 from http://economictimes.indiatimes.com/news/economy/agriculture/government-to-add-56800-tonnes-of-storage-capacity-for-onion/articleshow/52519598.cms

2. Madgavkar, A., and Krishnamurthy, S. (2014). Transforming the food chain byte by byte, December 29. Retrieved on January 23, 2016 from http://www.livemint.com/Opinion/ra1vcMcyCMWrOgukmgehhK/Transforming-the-food-chain-byte-by-byte.html

3. Seetharaman, G. (2016). How climate-smart agriculture is crucial for India's food security at a time of farm distress, May 08. Retrieved on May 31, 2016 from http://articles.economictimes.indiatimes.com/2016-05-08/news/72932519_1_rice-crop-tillage-taraori

4. Peisker, A. and Dalai, S. (2015). Data analytics for rural development, Indian Journal of Science & Technology, 8 (S4): 50-60.

5. Channe, H., Kothari, S. and, Kadam, D. (2015). Multidisciplinary model for smart agriculture using internet of things (IoT), Sensors, Cloud computing, Mobile computing and Big data analysis, Int. J. Computer Technology & Applications, Vol 6 (3),374-382

6. Tripathi, R., Shahid, M., Nayak, A.K., Raja, R., Panda, B.B., Mohanty, S., Thilgham K., and Kumar, A. (2012). Precision agriculture in India: Opportunities and challenges, October 29. Retrieved on January 24, 2016 from http://www.rkmp.co.in/content/precision-agriculture-in-india-opportunities-and-challenges

7. World Economic Forum. (2012). Big data, Big impact: New possibilities for international development. Retrieved on January 22, 2016 from http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

8. Addison, C. (2015). Data: the next revolution for agriculture in ACP countries, February 15. Retrieved on January 22, 2016 from https://brusselsbriefings.files.wordpress.com/2015/01/reader-bb40-data-revolution-eng_v2-low-res-proof2-2.pdf

9. Sylvester, G. (2013). Information and communication technologies for sustainable agriculture: Indicators from Asia and the pacific. Retrieved on January 22, 2016 from http://www.fao.org/docrep/019/i3557e/i3557e.pdf