Journal Home: www.biospub.com

Popular Article

Article: RT1099

Biotica Research Today Vol 4:9 2022

648/651

How to cite this article?
Narmadha *et al*., 2022. Big Data Analytics in Agriculture. Biotica Research Today 4(9):648-651.

Big Data Analytics in Agriculture

R. Narmadha*, P. Murali Arthanari and R. Arockia Infant Paul

Dept. of Agronomy, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu (641 003), India

Open Access

Corresponding Author
R. Narmadha
e-mail: narmadharaj97@gmail.com

Keywords
Agriculture, Big data, Data analysis, Precision agriculture

Abstract

In India, majority of the farmers are not getting the expected crop yield due to several reasons. In this situation, using multiple elements that influence production to identify crop adaptability and yield can improve crop quality and yield, resulting in higher economic growth and profitability. As a result, many farmers have begun to employ new technology and methods to improve their farming operations. Big data, machine learning and artificial intelligence (AI) can play a key role in this process. Big Data Analytics is a Data-Driven technology useful in generating significant productivity improvement in various industries by collecting, storing, managing, processing and analyzing various kinds of structured and unstructured data. Big data analytics are important to the core of various applications since data is the raw material which is fed as the input for processing. Volume, velocity, value, veracity and variety are the five V's which is considered as the characteristics of big data. Hadoop is the main framework for big data analysis which is open source software.

Introduction

The world population has been increasing rapidly as it becomes the most important reason for the food shortages. The United Nations Food and Agriculture Organization made an estimate of current world population is 7.7 billion and estimated it to hit 8.5 billion in the year 2030 and 9.7 billion in 2050. The most important problem is to take care of food production to the massive population by adequate food supply. Climate change, degradation of soil, pollutions, market fluctuation and government policies which add uncertainty to the food production and supply. The available resources like land, water and energy required for producing food to the increasing population is not adequate. It is therefore need to introduce the new technology and methods in the field of agriculture.

Digital agriculture knows to be a precision agriculture playing great role in farming areas where each and every movement of activities were controlled and monitored. New technology with data intensive approaches which helps to increase the productivity of crop by minimizing the environmental impact. The data collection is from the different sources such that sensors, photographs, satellite images *etc*. It will give the better understanding of crop and environmental behaviour which allow us to take decision on farm activities and management options. Internet of things and cloud computing which known to be smart farming which mean to be working in the real time situation. This helps the farmers to react to the sudden changes in the farm due to weather or disease incidence. The sensors and machineries across the field can keep track of temperature, soil condition, moisture and humidity in the form of data. The data collected from the various places of the farms for continuous time being leads

Article History
Received on: 20th September 2022
Revised on: 22nd September 2022
Accepted on: 23rd September 2022

E-mail: *bioticapublications@gmail.com*

to enormous volume of data. So, that the data collection was done using the sensors whereas to extract the meaningful information from the collected data it requires the systemic process. Thus the big data analysis becomes the great necessity. Big data in agriculture with the precision agriculture is new concept as it is emerging now with earl developmental stage. It not only helps in the farming activities but for the entire supply chain. Big data helps the farmers with real time condition of the farms, thus decision making for the farming activities becomes smart. "Big data is considered as a large collection of dataset which is having high velocity, volume and varieties that make difficulty in processing and managing by using traditional techniques and tools".

Big Data and Big Data Analytics

Big data is term for collection of data sets so large and complex that it becomes difficult to process using on hand database system tools or traditional data processing applications. These data are arranged in the various forms like structured or semi structured or unstructured data ranges in different sizes from terabytes to zettabytes. The diverse complex and large data sets require new form of integration to find the hidden values in data sets. The contribution of big data in our daily life which involves systems, sensors, mobile devices and ever digital devices are transmitting the data. The collected raw data will undergo the various processes like classification, processing and organization. Data analytics and storage tools facilitate to refine the data for decision making and future prediction analysis. The big data analytics can help to find the pattern, trends and discover the relationship from the collected data. It is core of different application where the data are fed as input for processing. Big data are in use with text mining, data mining and statistical analysis.

Types of Big Data Analytics

a) Descriptive Analytics

The set of historical data can be created in which it is said to be a preliminary stage. The data can be organized and patterns can be uncovered using the data mining techniques. Descriptive analytics can helps to find out trends and probabilities and gives information about what might happen in future (Riahi *et al.*, 2018). To discover the historical data the techniques used are data aggregation and data mining. In order to make the dataset more manageable the data will be gathered and sorted by the data aggregation. Data mining describes the next step of the analysis and involves a search of the data to identify patterns and meaning. Identified patterns are analyzed to discover the specific ways that learners interacted with the learning content and within the learning environment.

b) Diagnostic Analytics

It consists of asking the question: Why did it happen? Diagnostic analytics looks for the root cause of a problem. It is used to determine why something happened. This type attempts to find and understand the causes of events and behaviors (Riahi *et al.*, 2018).

c) Predictive Analytics

It is about the forecasting in which to predict the future it uses the past data. The current data were analysed using the data mining and artificial intelligence techniques to make the scenario what might happen in future (Riahi *et al.*, 2018). The software for predictive analytics has moved beyond the realm of statisticians and is becoming more affordable and accessible for different markets and industries, including the field of learning & development.

d) Prescriptive Analytics

It is dedicated to finding the right action to be taken. When historical data can be provided by the descriptive analytics and future prediction by the predictive analysis, the prescriptive analysis uses theses parameters to know the best solution (Riahi *et al.*, 2018). Prescriptive analytics is a statistical method used to generate recommendations and make decisions based on the computational findings of algorithmic models.

Characteristics of Big Data

Volume

It is one of the characteristics of the big data in which the data generated from various sources like social media, networks, business process, automated machines, sensors, human interactions, *etc*. on daily basis which is enormous in volume. The data generated which ranges from the size of terabytes to Exabytes to Zettabytes. This large amount of data is stored in the data warehouse.

Variety

The data which generated is in the various forms such that emails, PDFs, photos, videos, audios, *etc*.; whereas in the past times data were collected from spread sheets and databases. These forms are categorized under three types *viz*., structured, unstructured, and semi-structured data. *E.g.*, data in tabular format like excel will comes under structured data type, JSON, PNG *etc*., files comes under semi structured types of data and audio, videos *etc*., will comes under unstructured data types (Kumar and Menakadevi, 2018).

Velocity

The speed at which the data is created in real time is referred to as velocity. The ability to propagate knowledge, appreciation and innovation.

Veracity

The most challenging characteristics of big data. The accuracy of analysis depends upon the veracity of data. The bias, noises and abnormality in data is referred as the veracity.

Value

The value and potential derived from data.

Tools used in Big Data Analytics

There are many industry-standard BD analytics tools: Hadoop, MapReduce, HDFS, HIVE, HBase.

Hadoop

Hadoop is a framework for Big Data computing which has been released into the public domain as open source software, and so can freely be used by anyone. It consists of a number of modules all tailored for a different vital step of the Big Data process - from file storage (Hadoop File System - HDFS) to database (HBase) to carrying out data operations (Hadoop MapReduce). It has become so popular due to its power and flexibility that it has developed its own industry of retailers (selling tailored versions), support service providers and consultants.

Hadoop has two major layers namely,

• Storage layer (Hadoop Distributed File System), and

• Processing/ Computation layer (MapReduce)

Hadoop Distributed File System

The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware.

It has significant difference from the other existing distributed file system but also has many similarities. The low cost hardware with highly fault tolerant. It is highly suitable for application which is having the large datasets and provides high throughput access.

MapReduce

MapReduce is a computing procedure for working with large datasets, which was devised due to difficulty of reading and analysing big data using conventional computing methodologies. As its name suggest, it consists of two procedures - mapping and reducing. In mapping the available information are formatted according to the analysis, whereas in reducing the operation was performed as formatted. Hadoop framework also has two modules apart from two core components,

• **Hadoop Common** - These are Java libraries and utilities required by other Hadoop modules.

• **Hadoop YARN** - This is a framework for job scheduling and cluster resource management.

Precision Agriculture and Big Data Analytics

Approach of Information and Communication Technology in precision agriculture management would be an alternate method to utilize the new techniques in modern world. It requires data collection, analysis and processing the collected data. In Addition to this it requires some important techniques like Remote sensing, Satellite Navigation system, Geographical Information system, Automatic yield recording system, Automatic soil sensor, Variable rate technology, Advanced Farm Management.

The main asset of farming is soil where mapping will be controlled and measure the spatial variability. The nutrient, soil pH before and after production *etc*. can be monitored so that the future prediction can be done by utilizing Remote sensing, Satellite Navigation system and Geographical Information system which is recorded during field operation. The large amount of data which collected through the precision farming techniques can help the farmers to predict the future. It can be supported by the Big Data Analytics and ICT technology. Also, it can predict future or recommended decisions to farmers and vendors at the point of precision agriculture. It can help farmers to manage and increase yield productivity. The water management and irrigation automation become more accurate as per the soil humidity which improves the crop production. Web and mobile based applications visualize information from historical data, crop patterns and weather data. Big data analytics and ICT solutions helps the agriculture equipment companies and departments to perform analysis over agricultural growth and productivity, to support and identify future farming trends. The analysis supports policy makers, service providers, companies and government departments for deciding future varieties, pesticides and fertilizers.

Sawaitul *et al*. (2012) evaluated in back propagation algorithm model to predict the future weather. In weather forecasting the parameters like temperature, wind speed, relative humidity, rainfall will be recorded. By using the "Classification and Prediction of Future Weather by using Back Propagation Algorithm" the variation in one parameter says temperature it will predict what will be variation in other parameter like humidity and rainfall with respect to the temperature. Kamath *et al.* (2021) forecasted the crop yield using the random forest approach. In crop production prediction it involves huge amount of data which requires data mining.

Conclusion

The availability and openness of hardware and software, techniques, tools and methods for big data analysis, as well as the increasing availability of big data sources and datasets, shall encourage more initiatives, projects and

start-ups in the agricultural sector. This increasing availability of big data and big data analysis techniques, well described through common semantics and ontologies, together with adoption of open standards, have the potential to boost even more research and development towards smarter farming, addressing the big challenge of producing higher-quality food in a larger scale and in a more sustainable way, protecting the physical ecosystems and preserving the natural resources. Big data refers to the set of numerical data produced by the use of new technologies for personal or professional purposes. Big Data analytics is the process of examining these data in order to uncover hidden patterns, market trends, customer preferences and other useful information in order to make the right decisions. Big Data Analytics is a fast growing technology. It has been adopted by the most unexpected industries and became an industry on its own. But analysis of these data in the framework of the Big Data is a process that seems sometimes quite intrusive. Analytics is a data science. Big data takes care of the decision-making part while data analytics is the process of asking questions. Analytics tools are used when company needs to do a forecasting and wants to know what will happen in the future, while big data tools help to transform those forecasts into common language. More often, Big Data is considered as the successor to Business Intelligence.

References

Kamath, P., Patil, P., Shrilatha, S., Sowmya, S., 2021. Crop yield forecasting using data mining. *Global Transitions Proceedings* 2(2), 402-407.

Kumar, H., Menakadevi, T., 2018. A review on big data analytics in the field of agriculture. *International Journal of Latest Transactions in Engineering and Science* 1(4), 1-10.

Riahi, Y., Riahi, S., 2018. Big data and big data analytics: concepts, types and technologies. *International Journal of Research and Engineering* 5(9), 524-528.

Sawaitul, S.D., Wagh, K.P., Chatur, P.N., 2012. Classification and prediction of future weather by using back propagation algorithm-an approach. *International Journal of Emerging Technology and Advanced Engineering* 2(1), 110-113.