

Impacts of Big Data Analytics for Agriculture tools and techniques

Namrata Kumari^{1*}, Dr. S. M. Asif Ali²

¹ Research Scholar, Dept. of Computer Science & IT, Magadh University, Bodh Gaya, Bihar, India

Email- namratakrishnasinha@gmail.com

² Assit. Prof. & Head, Dept. of Physics Mirza Ghalib College Gaya, Bihar, India

Email- asif_ali_gaya@yahoo.com

Abstract - The present paper deals with “Impacts of Big Data Analytics for Agriculture tools and techniques.” A random sample 30 Various software platforms are developed to give information to farmers about new tools and techniques related to agriculture: (MySmartFarm, Awhere weather, Phenone, Farmlog, Datafloq, Farmeron) has been used by secondary source from Agriculture depart in the state of Bihar. The traditional tools and techniques (Big data) analysis massive amount of data. To store and analyze this type of data parallel computing and analyze paradigm is required. Big data analytic is used to weather changes and its impacts of agriculture in Bihar. From the big data analytic Agriculture framework is developed that identify disease based on symptoms similarity and recommend a solution based on high similarity and achieve their tools has been used. Then cleansing of data is done that is important information is extracted from unstructured redundant data and were normalization is done that is features are extracted from cleaned data. The data was used to analyze the agricultural tools and technique. It finds out disease name based on weather changes and its impacts of agriculture has been taken on past data. Result has been shows that data analytic in weather changes past 2018-2023 years that was

Keywords: Big data analytics, Tools and techniques, Weather, and Agriculture.

-----X-----

INTRODUCTION

New technology is transforming into every sectors. Today agriculture sectors are grown in intelligent world. With the rapid increase of data there arises the need of innovative technical and analytical strategies that are capable of handling the complex data structures. As a results with the increasing impact of world wide web (agriculture information websites, social media etc.) and internet of things (radio equipment's), agriculture has entered into the world of big data. Big data is a heterogeneous mixture of structured and unstructured data that is growing at an astonishing rate. Big data refers to the digital large-scale data that is difficult to manage and analyze using traditional software tools and technologies. Agriculture science researchers are discovering novel solutions for three major challenges of agriculture big data that are scalable infrastructure, management schemes and data mining analysis methods for large datasets. The focus of the paper is to develop recommendation system for crop disease control that will help researchers and agriculture officers in decision making from historical data with help of big data analytics. Discussed that if big data analytics is used in agriculture tools will not only be a great innovation in the history of human agriculture as well as a pioneering work in human history. Due to

technology the term data is replaced by transforming big data in agriculture fields. With the technological advancements and augmented growth of data agriculture data has entered the era of big data. Big data is a term used to depict augmented growth of data. Data may be in the form of file system or it may be in database. And that data can't be processed by traditional software techniques and databases and provide solution of agriculture crop diseases. With the help of big data agriculture analytics, researchers can easily make decision from historical data. It will be a great innovation and pioneering work in human history if big data analytics is used in agriculture. Agriculture data is increasing day by day at astonishing rate. The solution for this is to use big data analytics and for analysis of such type of data Hadoop and its tools are used. In the research work Apache Hadoop and Hive is used for solving the problems of agriculture big data analytics. Apache Hadoop is a platform used in distributed environment. system is developed that identify diseases and recommend solutions based on historical data and this framework will help researchers in decision making and it is easily understandable. The solution that is highly used for a particular disease has highest priority. For demonstration developed framework is

used to identify Paddy crop leaf blast disease and recommend a solution. Apache hadoop and its various tools are explained in next section.

In agriculture, most of the analytical methods are statistical based and are designed for analyzing single experimental dataset. Here one needs to rethink about the data analytics strategies to develop powerful tools with the capabilities that can analyze the big data in a better way. Researchers are trying to develop a large-scale data analytics tool using machine learning. The big data machine learning analytics tools Mahout and Rhadoop are already being used in healthcare sector, telecom industry, education system, banking sector, e-commerce etc. IBM introduced the term precision agriculture like healthcare last year. Precision agriculture analytics. The aim of the current research work is to develop a big data analytics recommendation framework to control crop disease that helps agriculture officers and researchers. The proposed work is easily understandable and helpful for recommending a solution based on evidence from historical data. There are various sectors in which big data analytics is used. Data in Agriculture sector is growing at rapid rate and it also enter in the era of big data. The presents an era on big data. In this survey paper, a brief overview on big data problems, including big data opportunities and challenges, current techniques and technologies. It also includes several potential techniques to solve the problem, including cloud computing, quantum computing and biological computing. Further it is describing predictive analysis on electronic health records using Hadoop and Hive. In this various data sources are used that are web pages, databases, flat files etc. Predict the healthcare benefits of different drugs and life style choice of patient. Risk factor of heart disease is identified based on LDL and HDL level of cholesterol. At ideal levels of diastolic and systolic patient's blood pressure is under control and have less risk of moving to next stage of hypertension.

Big data analytic: Modern agricultural tools refer to the involvement of technologies like GPS, sensors, A lot of devices to enhance the efficiency of planting, irrigation, harvesting etc. There are some ancient modern agriculture tools like tractors, cultivators also. In this article, we are focusing more on modern agricultural tools with the involvement of sophisticated methods. Some of the examples are drones, soil sensors, robotic weeders, crop monitoring apps etc.

In modern agricultural tools help farmers to lead a profitable and sustainable farming venture. Soil sensors are instruments used to measure soil moisture, pH, salinity conductivity, temperature. As in the process of agricultural cultivation, the soil condition needs to be periodically monitored to improve productivity. Based on the element the sensor measured from the soil, there are 5 types of sensors, namely, soil moisture sensor, soil temperature sensor, soil conductivity sensor, soil pH sensor

HYPOTHESIS

To analysed the big data tools and techniques for Agriculture.

METHODOLOGY

A random sample 30 Various software platforms are developed to give information to farmers About new tools and techniques related to agriculture: (MySmartFarm, Awhere weather, Phenone, Farmlog, Datafloq, Farmeron) has been used by secondary source from Agriculture depart in the state of Bihar. The primary motive of generation of results from the collection of data is to serve researchers by Agriculture.

It was not an easy task to develop a new framework identify disease and recommend solution based on symptoms similarity. These frameworks provide the solution based on historical data. Data for this framework is collected from various sources. This model basically works on recommendation system. The recommendation systems use the historical data or the knowledge of the product. Many e-commerce companies use recommendation system for sales. In the proposed model recommendation system is applied to agriculture domain.

Firstly, data is collected from various sources lab reports, agriculture websites etc. collected data is known as raw data because it contains irregularities and unwanted information. So data is unformatted and it needs formatting or confirmation. This data is stored on HDFS. Name Node of HDFS keeps track how your files are broken down into file blocks, which nodes store those blocks. clients communicate directly with Data Node to process the local files corresponding to the blocks. Data Acquisition: It consists of a typically sensing element that measures specific properties such as soil moisture, temperature, pH etc.

Signal Conversion: Then, the sensing element converts the measured data into electrical signals.

Signal transmission: The digital signals are then transmitted to a computer or controller through wired or wireless methods.

Data Processing: By receiving the data, the computer analyses and processes the data to extract useful information

Device control: Then by analyzing the results, the controller can automatically control systems such as irrigation, weather stations etc.

The benefits of big data mining are to provide- Precise irrigation control, Resource conservation, Improved crop yields and give real time data for better decision making.

REPORTS

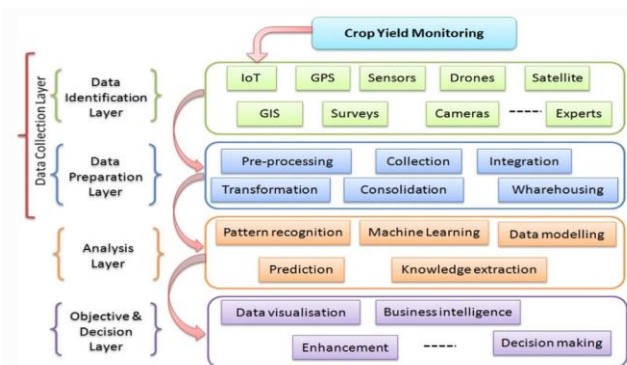
Agriculture department Reports provide information related to individual field of specific geographically area. The effects of high temperature on paddy and how to control it in specific district. Agriculture department (Reports) is very helpful for decision making of specific geographical area. In collected data is stored on Hadoop distributed file system in raw format. Row data contain large number of attributes and intolerance noise that make data meaningless. The solutions of the missing data problem are to delete incomplete data. Agriculture info towards websites act like mentor for farmers. These sites give information related to agricultural economic entity; commonly used pesticides etc. agriculture information websites provide information to farmers about which crop to plant where and when. And suggest solutions to various problems related to crops. by these sites farmers get knowledge about new techniques and tools.

Agriculture department reports: Using these reports decision making is easy for crops of particular area. These reports are important to provide information regarding particular field of a geographical areas.

RESULTS

Data that is collected from above sources is stored on Hadoop distributed file system in the form of text file. Collected data is unstructured and it contain irrelevant data. Firstly, unimportant data is removed and relevant data is extracted from collected data. Then features are selected and extracted from relevant data and save into text file on hive data warehouse. Hive is used to querying the data in distributed environment. Hive is open-source software tool used for data ware housing. To extract data out from Hadoop system Hive provides interface that is similar to SQL interface which is termed as HIVEQL HIVE query language.

In order to test the hypothesis that a big data analytic for weather changes of Agriculture on 2018-23 has been reported on fig-1.



Thrift server is used as an interface when client and server use different language. HiveQL extract data from hive data warehouse and save query results into text file that will store on HDFS. Now submit text file to distributed environment to identify crop disease name based on crop disease symptoms similarity. In this

process after splitting text file submitted to mapper to calculate pair-based symptoms similarity, pair-based similarity ignore spelling mistakes and word ordering this will increase efficiency of recommendation system.

After calculate similarity mapper create a pair, save into file and submitted to reducer, in this system disease name is key and similarity, solution and location are saved as values. Reducer calculates average similarity where disease name (key) is same and select high similarity disease. Now select a high similarity solution id from file that saves by mapper. Big Data analytics system architecture is depicted in fig-1. While this system is targeted specifically to crop yield management, it can be adapted to any data-driven application. This architecture implements faithfully what we have highlighted in the previous sections. In this section, we will focus on the data analysis layer of the architecture, moreover, we will pay attention to the data types and their sources, techniques of data acquisition, the learning algorithms.

CONCLUSION

With the advancements in technology agriculture data entered in the world of big data. Using Hadoop and HIVE tools big data analytical framework has been developed that will handle agriculture crop disease problems. Developed framework is useful for farmers and researchers for recommending a solution based on high similarity symptoms. The developed big data analytics framework is crop and location specific but the next step will be to develop a framework that will be dynamic in nature. Big data is not just characterised by the volume, but also by velocity, variety, and others. These are enough to challenge the existing data mining techniques, as trying to develop techniques to deal with large volumes of data (volume), various types of data attributes (variety or heterogeneity), and be able to analyse the new data as soon as it is collected (velocity) are extremely challenging tasks. Moreover, many other characteristics can be found in some big data-driven applications, these include veracity, value, viscosity, veracity, visualisation, etc. In this study, we added veracity, as the data, collected by various instruments and sensors, is of different quality, which creates a huge challenge to the data pre-processing task, and therefore its analysis. a systematic review of the potential use of the data mining process in crop production and management and highlighted serious gaps which can be considered in future studies. The majority of the current practices were dominated by statistical analyses and small machine learning systems. Nevertheless, despite all the advantages that can be gained from big data mining process are several other challenges and obstacles that need to be addressed, among them lack of data, lack of skills, and lack of maturity and standards so that it can be adopted and deployed quickly and easily.

REFERENCE

1. A, Konstan (2022) E-Commerce Recommendation Applications: Data Mining and Knowledge Discovery. Kluwer Academic Publishers. Manufactured in The Netherlands. pp.115- 153.
2. David B. Lobell. (2017), The use of satellite data for crop yield gap analysis.vol no. 143.
3. Ferstl et al., (2016), Time-hierarchical clustering and visualization of weather forecast ensembles, *IEEE transactions on visualization and computer graphics*, 23(1), 831–840.
4. Guocai Yang (2014), Agriculture Big Data: Research Status, Challenges And Countermeasures. Proceedings of Computer and Computing Technologies in Agriculture, China, 2014 September, 137-143.
5. Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, Guang-Zhong Yang. Big Data for Health. JULY, 2015; 19(4).
6. Kumpf et al., (2017), Visualizing confidence in cluster-based ensemble weather forecast analyses, *IEEE transactions on visualization and computer graphics*, 24(1), 109–119.
7. Marx V. (2013), Biology: The Big Challenges of Big Data. *Nature* 2013. 498(7453), 255-260.
8. Nayak et al., (2013), A survey on rainfall prediction using artificial neural network, *International Journal of Computer Applications*, 72(16).
9. Philip CL, (2022), Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Vol No. 275 ,314–347
10. Laney D. (2001), 3D Data Management: Controlling Data Volume, Velocity and Variety. Meta Group Inc Application Delivery Strategies; ADS (6), 1-4.
11. Wei (2017), Conceptual weather environmental forecasting system for identifying potential failure of under-construction structures during typhoons, *Journal of Wind Engineering and Industrial Aerodynamics*, 168, 48–59.
12. Xiao et al., (2018), Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data, *Remote sensing of environment*, 210, 48–64.
13. Xiaotong Lin (2022), Big Data Deep Learning Challenges and Perspective. *IEEE Access*. 2(1), Marx, VII. Biology: *The Big Challenges of Big Data*. p 255-260.

Corresponding Author
Namrata Kumari*

 Research Scholar, Dept. of Computer Science & IT,
 Magadh University, Bodh Gaya, Bihar, India

Email- namratakrishnasinha@gmail.com