

Big Data Analytics to Increase the Agricultural Yield by Using Machine Learning Approaches

V. Sudha¹, S. Mohan² and S. Arivalagan³

¹Research Scholar, Department of Computer and Information Science, Annamalai University, Tamil Nadu, India

^{2&3}Assistant Professor, Department of Computer Science Engineering, Annamalai University, Tamil Nadu, India

E-Mail: sudhaphd16@gmail.com, mohancseau@gmail.com, arivucseau@gmail.com

Abstract - Agriculture is the backbone of Indian economy. Big data are emerging précised and viable analytical tool in agricultural research field. This review paper facilitates the farmers in selecting the right crops and appropriate cropping pattern for a particular locality. A modern trend in the Agriculture domain has made people realize the importance of big data. It provides a methodology for facing challenges in agricultural production, by applying this Algorithm, using machine learning techniques. The different machine learning techniques survey is presented in this paper to realize enhanced monitory benefits in a particular area. A study of machine learning algorithms for big data Analytic is also done and presented in this paper.□

Keywords: Big Data Analytics, Machine Learning Algorithm, Precision Agriculture, Enhancement

I. INTRODUCTION

Big data is characterized by volume, variety, and quantity of transactions, events of historical data. Big data Analytics is the analytical dataset with some combination of data volume, data velocity and data variety. It includes complex analytics and complex data types. Big data is a term associated with the new dataset model and the technologies needed to alter in line with the new data source, which refers to big data analytics.

TABLE I TYPES OF BIG DATA AND ITS SOURCES

Data Types	Structured	Unstructured	Semi-Structured
Sources	Real time and large volumes such as retail, and finance	Images, Sensor data	Web Applications such as weblogs, email, WebPages
Formats	RDBMS, Data warehousing	User Generated Text	XML, CSV, JSON

II. LITERATURE REVIEW

Big data analytics application in agriculture has a huge potential. Agricultural systems that have been constructed and analyzed through the concept of big data analytics are quite limited compared to the available systems. GAO, "The impact of climate change on China's crop production" [1] have recommended and an implemented information system for agriculture is essential to sustainable development. They improved a statistical crop model using

the ancient yield and analyzed the impact of temperature and precipitation change on crop yields they analyzed. The statistical relationship between yield and climate change in the region. Using historical crop yield records simulations and predictions we studied the potential risk future climate change they post on the crop yields in the region. Climate model they analyzed the statistical relationship between yield and climate change for the regions.

P. S. Cornish *et al.*, [2] has presented the architecture of the big data open platform used for supporting Improving crop production for food security and improved livelihoods on crop options alternative cropping systems. Rain-fed vegetables are grown in the Kharif and then adapted by farmers to pre Kharif cropping and to the Rabi winter. They had some access to irrigation. How to reduce climate risk in rain-fed rice on risky medium Systems implemented by farmers needed for water resources, suggestion watershed development rice fallow with the safer climate responsive system. Small water harvesting structures may be needed for Rabi vegetable crops. The technology evaluated the process adoption of less risky cropping systems with greater water Productivity.□

K. Sabarina and N. Priya, [3] has presented an efficient strategy for crops Big data for the benefit of precision agricultural lowering data dimensionality Predictive analytics can be used to make the smartest decision in farming by collection real-time data analysis with streaming data. The data analysis efficiency challenge with the massive increase in the size of big data by applying a tensor-based feature radio model. Big data is one of the approaches to improve the data analysis performance of weather, soil, air quality crop maturity and labor costs and availability important role in precision agriculture in managing real-time data analysis streaming data. Real-time data gathered from different agricultural sources.□

M. R. Bendre, R. C. Thool, and V. R. Thool, [4] have recommended some implemented web-based information in Agriculture system, weather forecasting for future farming. Big data analytics in future process and challenges in prediction agriculture result using a programming model and elaborately distributed algorithm for data processing, agriculture applications provide a new insight to give advance weather. As decision improving yield productivity and avoid unnecessary cost related to harvesting. Big data

platform future application and challenges in precision agriculture result using a programming model and distributed algorithm for data processing and forecasting application of weather. As per this result, the model shows the temperature and rainfall in the region. Cropping pattern and Irrigation management are the solution to yield to increase the profit in Productivity.

J. W. Kruize, J. Wolfert, H. Scholten, C. N. Verdouw, A. Kassahun, and A. J. M. Beulens [5] Farm software ecosystem for smart farming, architecture monitoring, planning, and control of agricultural processes, smart farm reference architecture used to assess design and implement farm software ecosystems. Improve configuration of the different component. Farm software ecosystems they contribute to software development to enable smart farming. □

Ahrary and D. Ludena, [6] suggested a profile based architecture for precision agriculture to improve decision making in real time based on service. Internet of things to create integrated environment information research in agricultural and commercial field automated way in order to provide a richer experience to the user and to specific relevant information regarding the particular environment. Highly valuable for the nutrition based vegetable production and distribution system the multiple benefits of big data healthy food recommendation to the end user of the system different analytics to improve the system efficiency Technological solution to farmers without being expensive or difficult to use.

S. S. Reddy and C. S. Cork [7] for an integrated analytics approach towards agriculture. Big data analysis on density-based clustering algorithm advances in internet speed has enabled data on a global scale to be transmitted very high speeds the huge amount of data is being generated continuously as data streams from different real-world application. Clustering techniques for big data analysis, clustering is an unsupervised learning technique used to classify large data sets into clusters sharing similar qualities. Clustering data streams new challenges limited time, limited memory handling, noisy data and handling high dimensional data clustering.

S. Athmaja, M. Hanumanthappa, and V. Kavitha, [8] a survey of machine learning algorithms has presented effective strategies by for big data analytics. All over the world the agricultural peoples gained some advantages through the comparative knowledge from big data analysis, with machine learning algorithm by using huge data the agricultural peoples get some comparative knowledge and changes in regular agriculture.

J. Majumdar, S. Naraseyappa, and S. Ankalaki, “[9] have demonstrated and implemented information system Analysis of agriculture data using data mining techniques for agriculture planning intention is the accurate yield estimation for the numerous crops involved in the planning. Techniques are necessary to approach for accomplishing

and effective solution for this problem. Soil input levels, commodity price has made it all relevant for farmers to use information and get help to make a critical farming decision Existing crop, soil and climatic data and analyzing new data optimized the production and makes agriculture more resilient to climate changes. Input data to assess the best performance-yielding method.

P. Shah, D. Hiremath, and S. Chaudhary, [10] has demonstrated the big data analytics architecture for the agro advisory system and implemented an analytical engine on top of open source architectures. Agro advisory system big data analytics they can be used to improve the productivity in agriculture. Research is used to reduce the technological communities and information through recommendations and decision support system, propose an open source, cost-effective and scalable big data analytics architecture for an agree system crop yield prediction is built and implemented. Farmer to have an idea of yield based on ongoing weather parameters and to make a decision to grow that particular crop for alternate crop in case yield prediction is unfavorable. □

R. Kaur, R. Garg, and H. Aggarwal [11] increased some message for creating a frame worked solution for Identify crop disease based on previous symptoms and suggestions huge message cannot be stored in ordinary data so machine learning approaches are need for big data analysed framework. Java based Hadoop platform is used as a tool to collect this message. Developed a framework for generating recommendation solutions for diseases big data analytics framework is used to identify crop disease based on symptoms similarity and recommends a solution based on high similarity. Techniques are unable to store and analyse this amount of data to store and analyse this type of data parallel computing and analyse paradigm is required. Agriculture framework is develops Identify disease based on symptoms similar solution based on high similarity to achieve this objective Hadoop and Hive tools has been used, the data used is collected cleansed and normalized. Crop disease symptoms and propose a solution based on evidence from historical data is utilized to support the files. The kinds of files used are HDFS, Hive HiveQL and SQL.

H. Jain and R. Jain, [12] presented an overview of the current state of the art of web-based tools and technologies for processing large data set. Big data in weather forecasting Application and challenges what going to happen tomorrow and coming years. Applications could impact normal operations weather forecasting application and technical challenges. Agriculture got a new way to produce more crops on the same amount of land in an efficient and sustainable manner weather forecasting improving the strength and reliability, identify the optimal solutions more efficient and reliable applications to improve quality of business reduced risks and enhance profitability. □

P. Shah, [13] Spark based on agricultural information system domain at a higher velocity. Information is available in the form of weather and soil condition reports, water

resources, soil condition reports, field characteristics, and market conditions. Big data technology has huge potential to refer to this information and Produce Geospatial processing, remote sensing, and advance analytics algorithms, cloud resource and advance storage system. The system is proposed to collect, query, analyze and visualize heterogeneous and distributed data, including Geospatial data at scale using open source. Spark-based information management system for agricultural reduce the technological gap between agro users and information data. The implementation is done on big data open source architectures by developing various web-based analytical and visualization services.

S. Rajeswari, K. Suthendran, and K. Rajakumar, [14] have presented a conceptual architecture of the big data open platform used for supporting many sectors like agriculture. Big data, mining Techniques cloud-based big data analytics and IOT Technology performs an important role in the feasibility study of smart agriculture. Precision agricultural systems are estimated to play an essential role in improving agriculture activities.

Applications in a digital agriculture domain like monitoring the crop growth selection of the fertilizer, irrigation decision support system. Cloud-based big data analysis is analyzing the data requirements analysis the crops market and stock requirement for the crop. As the data is collected over the years for crop details and soil conditions this model provides Big Data analysis for best crop sequence next crop to be cultivated for better production, total crop production in the area of interest data analyses. □

II. AGRICULTURAL PROCESSES BY BIG DATA ANALYSIS

From Big data analysis farmers can identify the changes in rainfall, soil quality types of soil and forming times. So they are ready to face and get good results in harvesting. Machine learning technique increases the quality of the big data analysis. Now the crop yield predictions are automatically better even though the big data analysis is collected by machine learning techniques there may be chances for unstructured. Monitoring of big data analysis will give unstructured results only. So it is to be structured with proper clarity for forming as below.

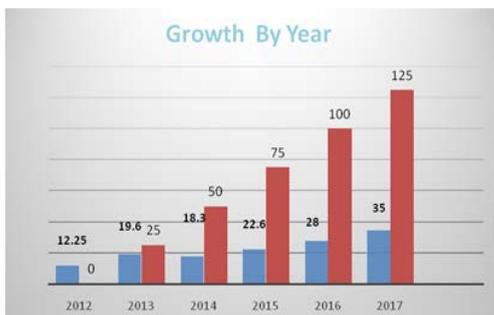


Fig.1 Big Data Processes Model

TABLE II ANALYZATION OF VARIABLE DATA

Characteristics	Big data	Traditional Data
Volume	Terabyte, Petabyte, Exabyte	Gigabyte
Variety	Structured, Semi-structured or Unstructured	Structured
Velocity	More rapidly	Per hour, day
Data Integration	Difficult	simple
Data access	Real-time □	Interactive
Source of data	Fully distributed	Centralized

III. MACHINE LEARNING TECHNIQUES

Machine learning research is developing vastly as efficient learning algorithms, which can make a prediction on large data. Machine learning algorithms were implemented for zip data sets. Machine learning is an approach used to improve scheduled models for prediction.

Machine learning tasks are grouped into three categories-supervised, unsupervised and reinforcement learning. Supervised machine learning requires training with labeled data. Each labeled training data consists of the input value and a desired target output value.

The supervised learning algorithm analyzes the training data and makes a conditional function. In the unsupervised machine learning technique, unknown approaches are drawn from unlabeled datasets, for example, cluster analysis.

Reinforcement learning allows a machine to learn its performance from the received connections with the external environment from a data processing point of view both supervised, unsupervised learning techniques are preferred for data analysis, and reinforcement technique is mostly preferred for decision-making problems.

IV. METHODOLOGY OF BIG DATA ANALYTICS

Big data Analysis with Machine Learning Approaches by Hadoop technique is applied to analyze the large volume of data. Clustering and K- Means Algorithm is used to identify the group of data and collaborate in one cluster.

Big data is used where the crops are dependent on the weather conditions by applying a different algorithm and tools. By using machine learning approaches under the platform Hadoop we can manage the environmental conditions, such like, soil types, soil testing, crop pattern, weather, crop rotation by proper parameter.

TABLE III CLASSIFICATION OF BIG DATA ANALYTICS

Domain Analysis	Characteristics	Techniques	Source of Data
Data Analytics	Structured records, Less volume and real-time.	Statistical Machine Learning, RDBMS, OLAP, Data Warehouse.	Business & Scientific Data.
Text Analytics	Language-Dependent Unstructured, semantic, Rich Textual□	Decision Tree, clustering, Naïve Bayes Neural Network, SVM.	Government Rules and regulations etc. Webpage, Corporate documents, Emails, Twitter, Face book.□
Web Analytics	Integration of Text and Hyperlink, Symbolic, Metadata	Classification and Clustering.	Various web pages.
Multimedia Analytics	Collection of Audio, video & Image	Multimedia annotation , Video, Audio summarization	User generated multimedia, surveillance, healthcare media, corporation produced Multimedia.
Mobile Analytics	Fragmented data person specific	Location Based Mining	Mobile Phone applications.

V. HADOOP TECHNIQUES

Big data analytics platform developed by Apache Hadoop is a parallel and distributed platform over the big data processing. Apache Hadoop and Hive are used for solving the problems of agriculture big data analysis the system recommends solutions based on historical data and this framework will help researchers in decision-making and it is easily understandable. Hadoop Technique helps for the data acquisition, management and visualization of applications in machine learning applications. It provides a Map Reduce algorithm for faster distributed and parallel processing, map-reduce is a programming model used for parallelizable problems across the large-scale data set. It is a Java based platform in Machine learning techniques to improve the big data analysis for the formers. It has highly compressed with complicated comparison data analysis for a long period observations in agriculture. It results a maximized analysis report to increase the convenience in deciding of harvesting to the authorities of the government. But now a day's Hadoop is advanced by Java.

Apache Mahout is mainly used in machine learning applications, and it aims to build an environment for quickly creating extensible performing machine learning applications. A hive is a tool used for data warehousing to extract data out from the Hadoop system. Hive provides the interface that is similar to the query language. Hive is used for querying data in a distributed environment. Hbase it is distributed and non-relational database management system. Hadoop Distributed File System (HDFS) is used for storing a large amount of data. Hardware may be a higher capacity Data node and name node name. It is used to predict weather forecasts to alert farmers and optimize crop efficiency to maximize the overall yield. Hadoop Distributed File System (HDFS) and its Contributions to agriculture technology also help farmers to decide when to plant and harvest crops.

VI. PROPOSED FRAMEWORK

Crop yield forecasting Intercrops Groundnut, Black Gram, Sunflower, and Green Dhal. Crop production they depend many factors other than the water to meet. Irrigation system,

Irrigation Management and water supply flexibility and reliability to Cropping Pattern to maximize the Productivity and Profitability to Farmer.□

VII. CONCLUSION

The paper has an overview of the method and algorithms used to recommend a solution based on a high comparison big data analytics framework is crop and location specific and nature. The principle of this study was to investigate preprocessing techniques make superior data for forecasting model. This model shows the best performance with different dataset the effective model coupled with data preprocessing Techniques for improving the accuracy of resources to create sophisticated and customizable working environment for users to improve the productivity in agriculture. Generally cultivation in agriculture is not easy, but periodical observations and proper experience only help in farming for initial farmers, when government starts to support farmers they create a data analysis for forming. Today highly compressed comparatively complicated algorithmic data analysis of machine learning techniques gives a booming output to the farmers even though they have multiple problems.

REFERENCES

- [1] Gao, "The impact of climate change on China's crop production: A CMP5 ensemble assessment," *1st Int. Conf. Agro-Geoinformatics, Agro-Geoinformatics 2012*, pp. 208–212, 2012.
- [2] P.S. Cornish *et al.*, "Improving crop production for food security and improved livelihoods on the East India Plateau II. Crop options, alternative cropping systems, and capacity building," *Agric. Syst.*, vol. 137, pp. 180–190, 2015.
- [3] K. Sabarina and N. Priya, "Lowering data dimensionality in big data for the benefit of precision agriculture," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 548–554, 2015.
- [4] M.R. Bendre, R.C. Thool, and V.R. Thool, "Big data in precision agriculture: Weather forecasting for future farming," *Proc. 2015 1st Int. Conf. Next Gener. Comput. Technol. NGCT 2015*, pp. 744–750, September 2016.
- [5] J.W. Kruize, J. Wolfert, H. Scholten, C.N. Verdouw, A. Kassahun, and A.J.M. Beulens, "A reference architecture for Farm Software Ecosystems," *Comput. Electron. Agric.*, vol. 125, pp. 12–28, 2016.
- [6] Ahrary and D. Ludena, "Research Studies on the Agricultural and Commercial Field," *Proc. - 2015 IIAI 4th Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2015*, pp. 669–673, 2016.

- [7] K.S.S. Reddy and C.S. Bindu, "A review on density-based clustering algorithms for big data analysis," *2017 Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud)*, pp. 123–130, 2017.
- [8] S. Athmaja, M. Hanumanthappa, and V. Kavitha, "A survey of machine learning algorithms for big data analytics," *2017 Int. Conf. Innov. Information, Embed. Commun. Syst.*, pp. 1-4, 2017.
- [9] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *J. Big Data*, vol. 4, no. 1, 2017.
- [10] P. Shah, D. Hiremath, and S. Chaudhary, "Big data analytics architecture for an agro advisory system," *Proc. - 23rd IEEE Int. Conf. High Perform. Comput. Work. HiPCW 2016*, pp. 43–49, 2017.
- [11] R. Kaur, R. Garg, and H. Aggarwal, "Big data analytics framework to identify crop disease and recommendation a solution," *Proc. Int. Conf. Inven. Comput. Technol. ICICT 2016*, vol. 2, 2017.
- [12] H. Jain and R. Jain, "Big data in weather forecasting: Applications and challenges," *2017 Int. Conf. Big Data Anal. Comput. Intell.* pp. 138–142, 2017.
- [13] P. Shah, "Towards Development of Spark Based Agricultural Information System including Geo-Spatial Data," pp. 3394–3399, 2017.
- [14] S. Rajeswari, K. Suthendran, and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," *Proc. 2017 Int. Conf. Intell. Comput. Control. I2C2 2017*, pp. 1–5, Jan. 2018.