Data Mining, Big Data Analytics and Their Applications in Agriculture Field

Abhisheik S
Computer Science and Engineering
Sri Ramakrishna Institue of Technology
Coimbatore.

Jaron J V
Electrical and Communication Engineering Sri
Ramakrishna Institue of Technology
Coimbatore.

Aiyswarya S
Information Technology,
Sri Ramakrishna Institue of Technology,
Coimbatore.

Surya Prakash S
Information Technology,
Sri Ramakrishna Institue Of Technology,
Coimbatore.

Abstract:- Data mining is the ever growing field which is used to extract hidden facts and unknown patterns from raw data, with the intent of turning this vast amount of data into useful information. Data mining is used in data science. Data mining automatically analysis data, classifies data and summarize data into useful information. Big data is the popular term in recent years. Big data is the term that describes a huge amount of structured, unstructured, and semi- structured data have been developed by various establishment around the world. This composite data is referred to as big data. Whereas big data analytics is the process of examining data sets in order to draw conclusion about the useful information they contain. In this paper we are going to discusses about data mining and big data analytics and their application in the agriculture field.

*Keywords : Big data,Agriculture,Data mining.*

I.INTRODUCTION:

The DATA MINING is the process of extracting hidden facts or data from the database. Data mining is used in data science. Data mining is practical in important sectors like health care, customer relationship management, marketing, ecommerce, fraudulent, insurance, banking etc. The 'BIG DATA' has been termed over the mid-1990s, it has become commonplace in our vernacular during the past decades[1]. **Big data** is an approach that used to analyze, systematically and extract information from it or otherwise it deals with Data sets. Big data challenges include capturing data, data storage, data analysis, sharing transfer, visualization, querying, updating, information privacy and data source[2]. The three key concepts of big data: volume, variety, and velocity. The first academic use of "big data" was in 1997 in the context of data visualization. The agriculture sector is volatile in nature due to dependencies on various parameter like weather forecasting, temperature control, soil fertility, supply chain[3]. The use of big data analytics along with live data as well as past data can be useful to predict accurate weather forecasting, temperature controlling, and soil nature. This will beneficial for formers to become self dependent for taking decision for seed growing pesticide controlling, irrigation controlling and monitoring the market prices for their crop yield.

II.CHARACTERISTIC OF BIG DATA:

Characteristic of big data can be classified into three forms;
1. Volume.
2. Variety.
3. Velocity.
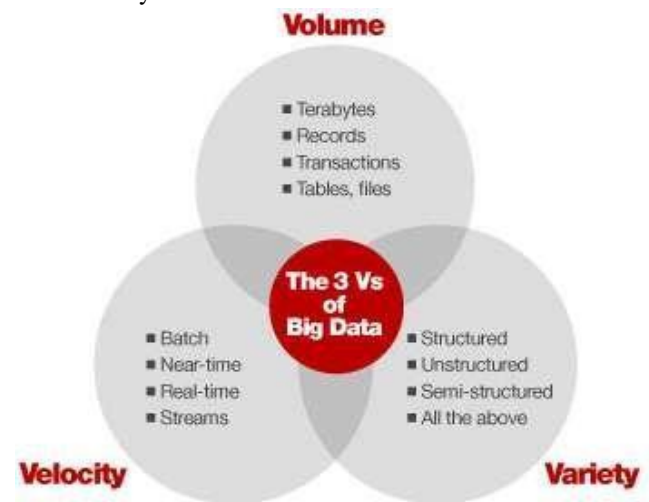
II.1.VOLUME:
Volume means storage capacity or amount of data. Collection of huge amount of data, records, transaction and tables in terabytes to peta bytes.

II.2.VARIETY:
Variety basically means the type of data. The types of data may be text format, click stream, web logs, images, video, animation, documents, sensor data and so on. These types of data are classified under structured, semi structured and unstructured data.

II.3.VELOCITY:

Velocity means generation of data from time to time. These data can be of real time data, near real time data, yearly data, monthly data, hourly data, historical data, weekly data and so on.

III.HADOOP-AN EFFICIENT TOOL FOR BIG DATA:

Hadoop is an open source, java based frame work developed in year 2006, manages by Apache software foundation. It is designed to store, process huge volume of data efficiently. Hadoop can be classified into two types:

1.Hadoop distributed file system(hdfs).
2.Map reduce.

The majority of this agricultural data is also unstructured, so it is a major challenge for this agriculture field to extract meaningful information regarding weather forecasting, temperature control, soil fertility, supply chain. Thus this Hadoop ecosystem can help the agriculture sector to manage this vast amount of data using big data analytics tool such as HDFS and map reduce.

IV.BIG DATA IN AGRICULTURE:

India's economical growth is depends on agriculture sectors as most of peoples in India depends on this sector[4]. Big data for agriculture sector, new term is introduced as Big data farming also called as precise farming. Precision agriculture is built upon the idea of 'smart farming', with the aim of addressing agricultural challenges in terms of growth, productivity, and provisions security and sustainability in response to climate change[5]. Smart farming takes advantage of the recent increase in the quantity, quality, and multiplicity of data generated from many sources including Geographic Information Systems (GIS), equipment sensors, climate and weather data, genomic information, as well as economic, social, and political data[6]. Much of this data is used to parameterize crop models that attempt to predict crop productivity (i.e. yield) in response to the environment, and ultimately global food security[7].

Remote sensing elements are used to gather the information. This gathered information is then visualized in an easy to understand format that is in structured data form[8]. One such example is the Normalized Difference Vegetation Index , which is a graphical indices used to assess the greenness of vegetation from remote sensing measurements. Data is also generated at the field level from various sensors attached to farm equipment, weather stations, field biosensors, and crowd sourced information from social media that, for example, report the incidence of natural disasters or pest infestations[9]. These data provide growers and researchers with spatial and sequential information about climate and local weather, soil conditions, crop quality, field biodiversity, and even

crop yields.

The recent innovation of high-throughput plant phenol typing relies on technology for automating oddity analysis in crops so that many more samples can be measured and analyzed than done manually[10]. These technology have contributed to the rapid collection of big amounts of phenol typical data that are more dependable and reproducible than those collected manually by teams of researchers[11]. They are used in reproduction programs to select plants with preferred characteristics. However, the longtime blockage in breeding programs has been the showing of crop genotypes under a variety of conditions to identify those that express behavior of interest. Often 100 to 1000 of different cultivars must be screened to find individuals with desired traits[12]. The concurrent growth and characterization of thousands of hereditary lines that can further developed for large-scale production . Three major methodologies are used in high- throughput phenol typing, and each produce at least gigabytes of data: remote sensing and imaging, laboratory analyses, and near-infrared reflectance spectroscopy. Currently, biologists, agronomists, and crop modelers are working together to develop tools to integrate data. Such multi-scale model has the possible to improve current crop models and provide more precise estimates about crop performance under untested ecological scenarios[13].

Technological advances in basic research will also increase the scale of data in agriculture field. These data help scientists to relate the original crop genetics to traits of interest, or provide insight about the molecular response of plants to genetic and to ecological changes[14]. The procreation programs has led to large increases in crop yield and biomass due to selected or engineered resistance to pests and other green stresses.

IV.1.USE AND APPLICATION IN BIG DATA:

Big data in agriculture field, often use in models geared towards improving crop yields and qualities in response to climate change[15. A recent study published in 2018 in the journal Remote Sensing of Environment utilized a variety of large scale data to forecast the optimal planting date for maize and soybean[16]. Sowing date is

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTICCT - 2019 Conference Proceedings

an important factor in formative crop yields globally, and later sowing dates often negatively affect yields due to increased heat stress and reduced moisture accessibility during reproductive and seed filling stages. This large-scale Multiple data sources have been used to study the impact of sowing date or crop yields. Improved plant Index from MODIS has been used for estimating the impact of sowing date on rice and soybean. However, a remote sensing data (EVI) along with fluorescence and radar data from novel satellite sensors, were used separately and in grouping to predict the sowing date for corn and soybean[17]. Thus from this study we came to know that by incorporating information about sowing-period temperature and crop fraction with satellite data led to higher model forecast exactness than using satellite data alone[18]. This use big datasets from many satellites, which increased the reliability of the model-predicted sowing estimates for multiple crops at regional and global scale[19]. Thus big data used in agricultural field to improve the crop productivity.

V. CONCLUSION:

Thus Data mining and Big data generate, collect and analysis in agriculture has improved crop and field management strategies, often resulting in improved yields. The development of new analytical tools and models to integrate, meaningfully[20]. the terabytes of data generated annually, may lead to the design of crop videotapes that will maximize yields under different environmental conditions. The big data generated through precision agriculture technology may help move toward prescriptive agriculture that is dynamic and efficient[21]. Both for-profit and non- profit organizations are developing tools to forecast plant-environment interactions to facilitate early and targeted intervention that will improve overall crop productivity, and ideally, global food security[22]. This will help the farmer to become self dependent and it also help them to take decision for seed growing, pesticide controlling, irrigation controlling.

REFERENCE:

[1] Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 19, 52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

[2] Budak, H., Hussain, B., Khan, Z., Ozturk, N.Z., Ullah, N., 2015. From genetics to functional genomics: improvement in drought signaling and tolerance in wheat. Front. Plant Sci. 6 https://doi.org/10.3389/fpls.2015.01012.

[3] Cox, M., Ellsworth, D., 1997. Managing big data for scientific visualization. ACM Siggraph 97, 21–38.Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V.R., Tsou, A., Weingart, S., Sugimoto, C.R., 2015. Big data, bigger dilemmas: a critical review. J. Assoc. Inf.Sci.Technol.66,1523–1545. https://doi.org/10.1002/asi.23294.

[4] Frank, T., Meuleye, B.S., Miller, A., Shu, Q.Y., Engel,

[5] K.H., 2007. Metabolite profiling of two low phytic acid (lpa) rice mutants. J. Agric. FoodChem.55,11011–11019. https://doi.org/10.1021/jf0723559.

[5] Gandomi, A., Haider, M., 2015. Beyond the hype: big data concepts, methods, and analytics. Int.J.Inf.Manag.35,137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

[6] Guan, K., Medvigy, D., Wood, E.F., Caylor, K.K., Li, S., Jeong, S.-J., 2014. Deriving vegetation phenological time and trajectory information over africa using SEVIRI daily LAI. IEEE Trans. Geoscience Remote Sens. 52, 1113–

[7] 1130. https://doi.org/10.1109/TGRS.2013.2247611. [7].Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Ullah Khan, S., 2015. The rise of "big data" on cloud computing: review and open research issues. Inf. Syst. 47, 98–115.

[8] https://doi.org/10.1016/j.is.2014.07.006.

[9] Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron.Agric.143,23–37. https://doi.org/10.1016/j.compag.2017.09.037.

[10] Kembhavi, A., 2018. Big data in

[11] astronomy and beyond. In: Munshi, U.M., Verma, N. (Eds.), Data Science Landscape: Towards Research Standards and Protocols. Springer, pp.59–66. https://doi.org/10.1007/978-981-10-7515_4.

[12] Kramer, M.G., Redenbaugh, K., 1994. Commercialization of a tomato with an antisense polygalacturonase gene: the FLAVR SAVRTM

[13] tomatostory.Euphytica79,293–297.

[14] https://doi.org/10.1007/BF00022530.

[15] Lius, S., Manshardt, R.M., Fitch, M.M.M., Slightom, J.L., Sanford, J.C., Gonsalves, D., 1997. Pathogen-derived resistance provides papaya with effective protection against papaya ringspot virus. Mol. Breed. 3, 161–168. https://doi.org/10.1023/A:1009614508659.

[16] Lobell, D.B., Hammer, G.L., Chenu,

[17] K., Zheng, B., Mclean, G., Chapman, S.C., 2015. The shifting influence of drought and heat stress for crops in northeast Australia. Glob. ChangeBiol. 21,

[18] 4115–4127. https://doi.org/10.1111/gcb.13022.

[19] Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., Hammer, G.L., 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest.Science344,516–519. https://doi.org/10.1126/science.1251423.

[20] Marshall-Colón, A., Long, S.P., Allen, D.K., Allen, G., Beard, D.A., Benes, B., von Caemmerer, S., Christensen, A.J., Cox, D.J., Hart, J.C., Hirst, P.M., Kannan, K., Katz, D.S.,Lynch, J.P., Millar, A.J., Panneerselvam, B., Price, N.D., Prusinkiewicz, P., Raila, D., Shekar, R.G., Shrivastava, S., Shukla, D., Srinivasan, V., Stitt, M., Turk, M.J., Voit, E.O.,Wang, Y., Yin, X., Zhu, X. - G., 2017. Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform.Front.PlantSci.8https://doi.org/10.3389/fpl s.2017.00786.

[21] Padgette, S.R., Kolacz, K.H., Delannay, X., Re, D.B., LaVallee, B.J., Tinius, C.N., Rhodes, W.K., Otero, Y.I., Barry, G.F., Eichholtz, D.A., Peschke, V.M., Nida, D.L., Taylor, N.B.,Kishore, G.M., 1995. Development, identification, and characterization of a glyphosate-tolerant soybean line.CropSci.35,1451–1461. https://doi.org/10.2135/cropsci1995.0011183X0035

[22] 00050032x.

[23] Perlak, F.J., Stone, T.B., Muskopf, Y.M., Petersen, L.J., Parker, G.B., McPherson, S.A., Wyman, J., Love, S., Reed, G., Biever, D., Fischhoff, D.A., 1993. Genetically improved potatoes: protection from damage by Colorado potato beetles. Plant Mol. Biol. 22, 313–321. https://doi.org/10.1007/BF00014938.

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTICCT - 2019 Conference Proceedings

[24] Röhlig, R.M., Eder, J., Engel, K.H.,

[25] 2009. Metabolite profiling of maize grain: differentiation due to genetics and environment. Metabolomics5,459–477. https://doi.org/10.1007/s11306-009-0171-5.

[26] Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T., Thomas, A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the

[27] 21st century in a global gridded crop model

[28] intercomparison. Proc.Natl. Acad. Sci. 111, 3268–

[29] 3273. https://doi.org/10.1073/pnas.1222463110. [19].Sánchez, N., González-Zamora, Á.,

[30] Martínez-Fernández, J., Piles, M., Pablos, M., 2018. Integrated remote sensing approach to global agricultural drought monitoring. Agric. For.Meteorology 259, 141–153. https://doi.org/10.1016/j.agrformet.2018.04.022.

[31] Urban, D., Guan, K., Jain, M., 2018.

[32] Estimating sowing dates from satellite data over the U.S. Midwest: a comparison of multiple sensors and metrics. Remote Sens. Environ. 211,400–412. https://doi.org/10.1016/j.rse.2018.03.039.

[33] Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M.A., San Vicente, F., Olsen, M., Buckler, E., Jannink, J.L., Prasanna, B.M., Crossa, J.,

[34] 2015.Genomic prediction in biparental tropical

[35] maize populations in water-stressed and well- watered environments using low-density and GBS SNPs.Heredity114,291–299. https://doi.org/10.1038/hdy.2014.99.