

Survey on Data Mining Techniques for Agriculture

Harshavardhan.J. R¹Lathashree. R² S. Lekhashree³Akash.J. V⁴ Mohammed Rashid⁵

^{1,2,3,4,5}Department of Computer Science &Engineering

^{1,2,3,4,5}Visvesvaraya Technological University,Belagavi,Karnataka, India

Abstract—Data mining in agriculture is a very recent research topic nowadays. It consists of the application of data mining techniques in agriculture. Recent technologies are nowadays, able to provide a lot of information on agricultural-related activities, which can then be analyzed in order to find important information. It is also referred to as precision agriculture. In agriculture sector where farmers and agri businesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. In Big data agriculture has been a recent topic.. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. This paper focuses on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using data mining techniques like Sammon's Mapping, Self Organizing Maps (SOM), Map reducing techniques, multiple Linear Regression., NoSQL, and Mining is the large amount of existing crop, soil and climatic data, and analysing new non-experimental data. Optimizes the production and makes agriculture more resilient to climatic change.

Keywords—Multiple Linear Regression, Sammon's Mapping, Self Organizing Maps (SOM), Map reducing techniques, Sammon's Mapping, Self Organizing Maps (SOM)

I. INTRODUCTION

In the recent years, the huge volume of real time data in the agricultural sector and its need for an efficient and effective processing, stimulate the use of novel technologies and platform to acquire, store, process, analyze and visualize large data sets for future predictions and decision making. Big Data is an evolving term given to a wide area of data-intensive technologies in which the datasets are extremely large that dealing with them become more challenging than how it was before. Due to the critical challenges facing the agriculture sector farmers feel more forced to adopt intensive farming practices and sustainable agricultural ones, in order to increase both economic and environmental costs.

II. LITERATURE REVIEW

In paper[1] prescient examination that make investigation of various sorts of parameters that gives the examination. Cultivating basic leadership rely upon numerous parameters like soil condition, climate condition and for showing signs of improvement promoting and income likewise need to dissect the economic situate ions.

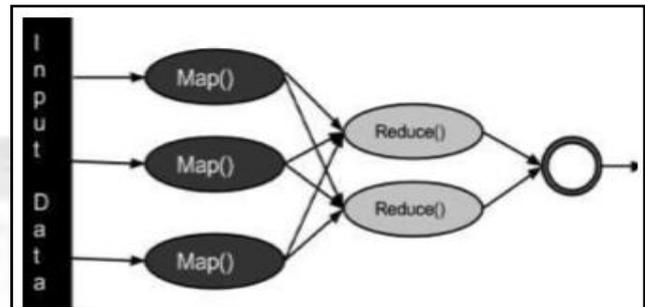


Fig. 1: Mapreduce Model

The proposed answer for this sort of issue utilize enormous information for conveyed figuring with delineate calculation and in addition business expository application like pentaho BI that give 3D information representation with various parameters like climate, soil investigation and economic situation with visual diagram and charts by utilizing this we can settle on choice of which compose crops appropriate for current circumstance.

In paper[5] is proposing a model coordinating farmers,agriculture agencies,agriculture expansion officers service of horticulture and agribusiness bolster institution. In the proposed show agriculturists will be given every one of the information they require to settle on choices on what product to develop contingent upon climate condition, water table content, the soil write, advertise availability among others. Knowledge construct contains areas with respect to cultivating procedures offering information, crop data and geospatial data .Farming strategies incorporate the best practice in cultivating to succeed most extreme yield and new cultivating systems that activate cultivating methods. Selling data incorporates the insights about venders, dealers, warehouses, funds, credits, dedicated website, call centre, E-learning and others. Aim of the model is to diminish labor and to build the ranchers potential to awesome degree.

In paper[6] through crop yield prediction system better planning and decisions can be chalked out for enhancing the yield

- 1) Input Most of the research papers that were studied have considered some climatic parameters like temperature, humidity, rainfall. Some agronomical parameters like soil, nutrient contents like N, P, K, and pesticides etc. The values of these variables have been taken as input.
- 2) Preprocessing (Noise Removal) for the successful application of data mining a huge set of dataset is required. The data which is acquired from various resources are sometime in raw form. It may contain some incomplete, redundant, inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized
- 3) Feature Extraction (Attribute Selection) This step aims at identifying and using most relevant attribute from the dataset. Through this process irrelevant and redundant information is removed for the application of classifier

4) Output The output is the crop yield prediction per acre with some recommendation. A broad outline of the crop prediction approach is shown in figure 3.

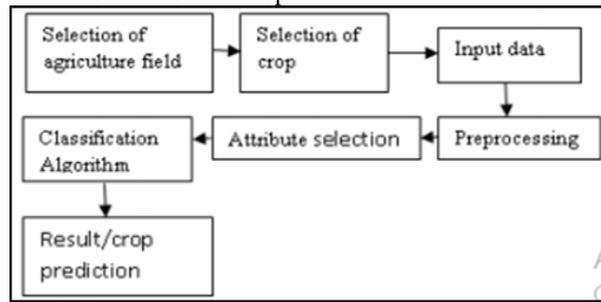


Fig. 3: crop Prediction system

| Algorithm/models used | Crop type | Accuracy/Recommendation |
|---|-----------------------------|--|
| Multiple Linear Regression | Rice yield | 90%-95% |
| Decision tree analysis and ID3 | Soybean | The rules formed from the decision tree are helpful in predicting the conditions responsible for the high or low soybean crop productivity under given climatic parameters |
| Support Vector Regression model | For any crop | The results show that support vector regression can serve as a better reference model for yield prediction. It is computationally less demanding. |
| Three models used APAR,SEBAL, Carnegie Institution Stanford model | Wheat,rice,sugarcane,cotton | Successful for wheat,rice ,sugarcane but not successful for cotton.The proposed technology can significantly contribute to quantitatively describing yield variations across the Indus Basin |
| Neural Network | Corn yield | 95% |
| C4.5 algorithm and decision tree | Soya bean,pady,maize | For soyabean=87% For paddy=85% For maize=76% |
| Harmonic Analysis of NDVI Time-Series algorithm | sugarcane | 86.5% |
| Gaussian Processes | Wheat yield | Wheat yield is expected to increase with an increase in temperature but there can be anincreasing underestimation error inPredicting the wheat yields |
| LADTre | Rice | 100% |

Table 1: Various Big Data Techniques

In This Paper [7] is talking about two different methods taken into consideration:

- 1) Self-Organizing Maps (SOM): It is a type of artificial neural network to produce a low dimensional representation of the data. One of the most popular neural network models and a type of unsupervised learning is Self-Organizing Maps (SOMs). It was developed by professor Kohonen. No human intervention is needed during the learning and very less knowledge about the input characteristics is required in SOM. Kohonen's SOM is known as topology-preserving map because in these topological structure is imposed on network nodes The SOM defines an ordered mapping in which a given set of data items is projected onto a regular two-dimensional grid SOM follows unsupervised learning. SOM provides mapping from high dimensional space onto a plane. Relative distance between the points is preserved using SOM mapping. There are four major components of self-organization process they are Initialization, Competition, Cooperation and Adaptation.
- 2) Sammon's mapping: Sammon's Mapping is part of Multidimensional Scaling technique. The most important goals in visualizing data is to sense how far or near points are from each other. In many cases, a

scatter plotting of those data might not be useful. For some kinds of data Euclidean Distance technique is useful. MDS is set of these kinds of technique to address these type of problem. It allows to visualize how near points are to each other from different types of distances or dissimilarity and can produce representation of data in small number of dimension. Sammons mapping is a multidimensional mapping technique used for transforming dataset with high-dimensional space to low-dimension. It is used for mapping d dimension dataset to m dimension non-linear sapce.It is often used for visualizing highdimension dataset to 2D or 3D. It can also use to map dataset with low-dimension space i.e output is not restricted to 2D.The distance -preserving can be important when user/one wants to use classifiers sensitive these distances like nearest-neighbor classifiers. Sammonsprojection is a multi-dimensional scaling method which is a widely used variant for non-linear projection. The sammon's mapping tries to match the low dimensional projections of data items with their original distance.

In paper [2] proposed BDA design is partitioned into three segments; Establish enormous information pipeline, Agro Advisory System and Data store. There are

four sorts of client connect with the framework; 1) framework designer, 2) information researcher, 3) area master/agrarian researcher, and 4) clients or agriculturists. Proposed engineering A.Perish huge information pipeline The greatest test in BDA application advancement is to bring information energetically. Choice of best in class enormous information innovation for huge information administration is the fundamental issue for information planners/Engineers. The primary part of framework engineer is to set up a major information pipeline. A major information pipeline in the proposed design is primarily partitioned into five phases; 1) information extraction, 2) crude information stockpiling, 3) information preprocessing and reconciliation, 4) systematic motor, and 5) information perception.

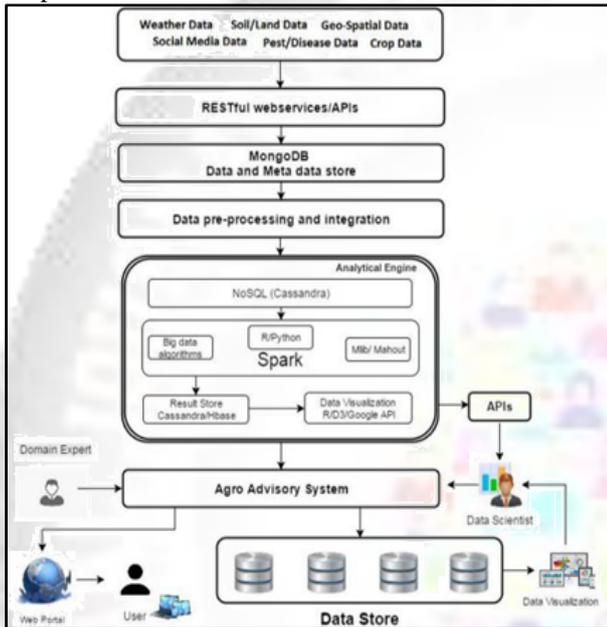


Fig. 4: Big Data ArchitectureFigure

In paper[3], Two key holes confronting the farming area have been recognized. They incorporate the failure to meet residential nourishment necessities and the powerlessness to send out at quality levels required for showcase achievement. We visualize that these issues can be unraveled by utilizing web of things (IoT) and information examination (DA). In this paper, the utilization of IoT innovations and DA in horticulture are talked about. The advantages and difficulties of sending the IoT and DA are displayed. At long last, strategies that can be received towards taking care of the issues confronting Nigeria's agribusiness area are proposed. World populace is evaluated to be around 9.7 billion of every 2050, all things considered the interest for more nourishment for the developing populace ought to be one of the key enthusiasm for creating nations and the world everywhere. IoT can be utilized to adjust the extraordinary climate conditions, the changing atmosphere and threatening condition (substance organizations, surge, dry season and soil disintegration) with the cultivating practice in the country. In this area, we examine the few IoT organization situation in farming and their advantages.

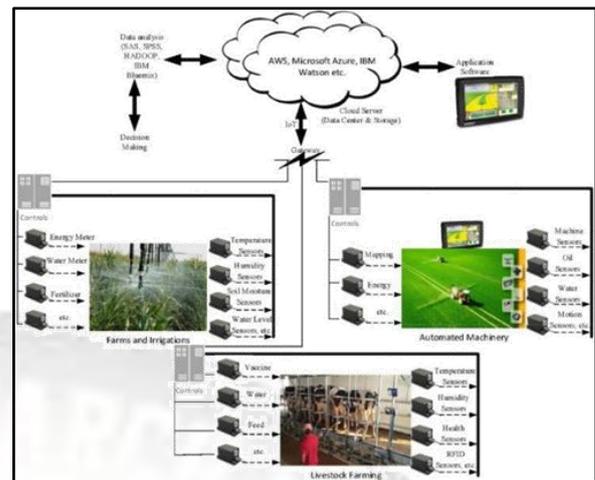


Fig. 5: IoT and Big Data in Agriculture

In paper[4], Big Data applications are utilized to remove profitable data from extensive informational collections produced from various sources, for example, web-based social networking sites, logs, computerized recordings and pictures, and sensor systems. Organizations in the farming area utilize chronicled atmosphere and sensor arrange information to catch and consolidate data that can help with the basic leadership process. At the point when the volume of sensor information expands, the abilities of a solitary server to store, process and deal with the information end up wasteful. Huge information advances are identified with dispersed figuring and huge information preparing ideal models. One structure that gives this worldview is Hadoop1. Bolstered by Apache Software Foundation, this structure has two essential sections: the Hadoop Distributed File System (HDFS) and Map Reduce programming perspective. The two sections outfit a circumstance with uprightness, execution, openness, and adaptability. Guide Reduce getting ready model has Job Tracker as expert and Task Trackers as slaves. Employment Tracker is responsible for enduring and administering occupations set up together by clients. It designates and screens plot reduce endeavors to be taken care of by the Task Trackers. The guide work contains a plan of data and has the ability to transform it into a summary of key/regard sets. The reduce work takes the came to fruition rundown from the guide work and influences a respect report containing the decreased once-over of key/regard sets.

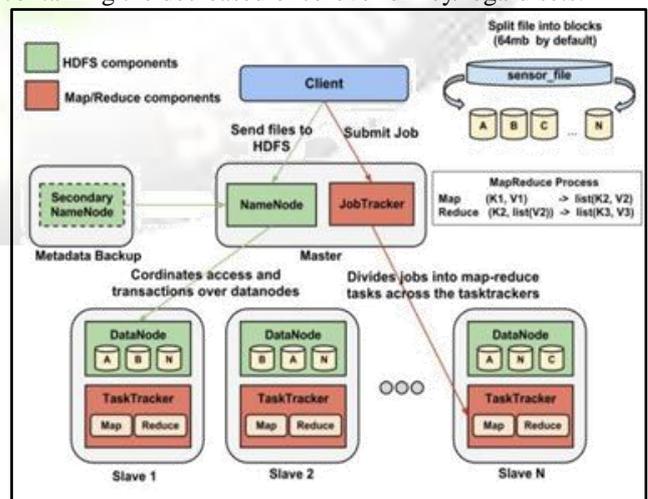


Fig. 6: Hadoop Architecture

III. CONCLUSION

From all the above discussed papers, it is examined about the profile based way to deal with oversee farming information inside a distributed computing design. Our approach will direct the enormous information suppliers and the different agrarian performing artists to distinguish and select the best administrations adjusted to their particular needs. By utilizing our approach, the rural performing artists can without much of a stretch incorporate the universe of enormous information clients, and advantage from the upsides of Big Data advances. All things considered, the work done in this paper is centered around the Big Data benefit arrangement and the profile benefit recognizable proof and determination. In a future work, our approach will be stretched out by building up an application to computerize the procedures of determination and piece of Big Data administrations.

REFERENCES

- [1] Mukesh Kumar and Prof. Mayuranagar “Big Data analytics in agriculture and distribution channel” Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication(ICCMC)
- [2] Purnima Shah, Deepak Hiremath and Sanjay Chaudhary “Big Data Analytics Architecture for Agro Advisory System ”IEEE 23rd International Conference on High Performance Computing Workshops,2016
- [3] Olakunle Elijah, IgbafeOrikumhi , Tharek Abdul Rahman ,Suleiman AliyuBabale and Stella IfeomaOrakwue “Enabling Smart Agriculture in Nigeria: Application of IoT and Data Analytics” IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON),2017
- [4] Rosangela de Fátima Pereira*, Marcelo Risse de Andrade*, ArturCarvalhoZucchi*, Karen Langona*, Walter Akio Goya* Nelson Mimura Gonzalez*, Tereza Cristina MeloBrito de Carvalho*, Jan-EriMångs† and AzimehSefidcon “Distributed processing from large scale sensor network using Hadoop” ,IEEE Interenational congress, ,2013
- [5] Mohanraj I and J.Naren “An Architectural Framework for E-Agricultural System” InternationalConference on Computing for Sustainable Global Development(INDIACom), 2016
- [6] YogeshGandge and Sandhya”A Study on various Data Mining Techniques for Crop Yield Prediction” International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2017
- [7] YashSanghvi, Harsh Gupta, HarmishDoshi, DivyaKoli, AmoghAnsh ,Umang Gupta” Comparison of Self Organizing Maps and Sammon’s Mapping on agricultural datasets for precision agriculture” IEEE Sponsored 2nd International Conference on Innovations in Information ,Embedded and Communication systems (ICIIECS)2015