International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

Multidisciplinary Approaches using Internet-of-Things (IoT) and Big-Data Analytics in Agriculture to produce high crop yield

1.T.Giri Babu,Research Scholar ,2. Dr.G.Anjan Babu,Professo
1,2.Dept. of Computer Science, S.V University ,Tirupathi, Andhra Pradesh, India
1.telugu.pplgiri@gmail.com,2. gabsvu@gmail.com

Abstract:
India is an agricultural powerhouse to the world. At present, India ranks second position in farm output all over the world. Despite the fact that agriculture accounts for as much as a major part of the Indian economy and employs an estimated 63 percent of the labour force, it is considered highly inefficient, wasteful, and incapable of solving the hunger and nutrition problems. Big data is playing an important role in modern agriculture. Major design techniques in farm equipment, including the use of sensor and smart technologies, now make it possible to collect vast amounts of digital information about crop applications, crop data, and other farm operations. With the advent of Internet of Things, it is easy to receive highly accurate, real time information about exact agricultural processes like sowing of seeds, harvesting, soil information, disease information and so on. Farmers can also collect real time information like the quality of soil, weather conditions and more.

This paper presents an enhanced automated techniques based on hadoop framework for efficient and scalable data analysing and forecasting system that are useful in the agricultural and allied sectors. This paper gives a brief idea about the importance of IOT in agriculture sector.

Index terms:Hadoop; big data; crop yield; HDFS

I.INTRODUCTION

The advent of Internet of Things (IoT) and Big Data Analytics into agriculture has been made in very good time because the world population is expected to increase to 10 billion by 2050. It would have been hard to feed all those people when there are inefficient farming methods. In order to keep up with the food requirements, it is required that a new and innovative method of farming methods and techniques should be introduced. Some of the methods include, reduce the wastage of water and power while harvesting, knowledge of what kind of seeds to plant and what kind of fertilizers to use and when, deciding harvesting time etc.

Despite progress in agriculture sector, several problems have continued to frustrate for decades. It is estimated that 30 percent of the total agricultural output is lost due to inefficiencies in harvesting, planting, transport, and storage.
Agriculture sector challenges will be important to India's overall development. All measures to increase productivity will need to encourage, amongst them: increasing crop yields, reducing water and power wastage, and developing accurate methods to reduce marketing costs. There exist lot of methods and proposed models for the prediction of crop status decisions with innovative ways of analysing and classifying data sets.

2.Related work

Agricultural decisions are made within a collective and perfect environment. Perhaps half of the world's farm households are part of collectivized or communal agriculture, and yet these farmers, like their private counterparts, must still make many decisions that are not made by agricultural officers. Much of farmers' daily work is done at their own initiative, and the incentives that encouraged them to work in a timely and careful fashion strongly affect the quality and quantity of agricultural output. In both personal and group agricultural communities, the decision making environment is conditioned by incentives to work. Identifying the factors that influence the size and quality of agricultural output is impossible without an understanding of the decision making environment of the farmers.

Key Decisions made in each crop or season

1) Planning

This is the first phase in the crop season. In this step, the basic requirements that is helpful for the crop harvesting is

important criteria. The basic criteria involved are

a) Equipment selection : The type of equipments we are going to utilize in crop harvesting.

b) Result analysis : The prediction of field output and market trends that are to be faced

c) Seed selection : The type of seeds and variety of seeds to be sowed in the fields.

d) Crop selection : The type of crop to be selected according to the climate and other conditions.

Here, the steps to follow and the decisions are to be taken in an appropriate manner are given.

2) Pre-planting
In Pre-planting phase, the steps that are to be practised before sowing of seeds is to be analysed.
a) Fertility management : The type of fertilizer to be used while crop harvesting.
b) Water management : The usage of water in the agriculture must be reduced.

3) Planting
This is the main step in the crop harvesting. In this step the actual planting and pre-planting can be done. The main steps are a) Planting logistics b) Planting practices

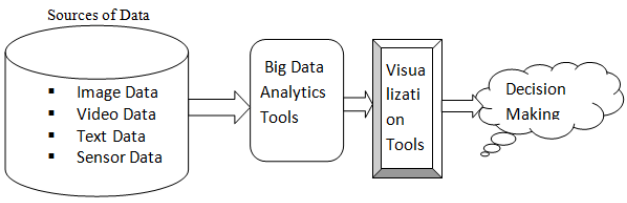
4) In season
This is the middle stage of the crop harvesting. The inputs that are to be added to the crops are to be identified. The main steps are a) Scouting b) Inputs c) Fertility

5) Harvesting
This is the last stage of the crop harvesting. The marketing of the crops are also very important in the agriculture sector.
a) Harvest logistic b) Harvest marketing c) Grain marketing

III.Proposed System

IoT based agriculture production system involves the monitoring system to analyze crop environment and also provides method to enhance the efficiency of decision making by analyzing crop data, forecast agriculture production using IoT sensors. By using the sensors, we can get the accurate information about species, fertilizing, soil and much more is collected in one database. Sensors in agriculture play an important role today. In agriculture, the need for increasing the production and at the same time the efforts for decreasing the environmental impact and for reducing costs make the sensor systems the best application tool. The use of sensors helps to exploit all available resources appropriately and to apply hazardous products considerably. When nutrients in the soil, humidity, solar radiation, density of weeds and all factors affecting the agricultural production are known, this gets better and the use of chemical products such as fertilizers, fungicides and other pollution products can be reduced moderately.

Architecture of Proposed system:



The above Figure shows the architecture of the proposed system in which Map Reduce in the use of Hadoop to monitor and analyzing the big data collected from various forms like rainfall data, sensor data, video and image data, then result has to be presented to Big Data analytics to make accurate decisions, encourage farmers to reduce expenses in agriculture, and to protect the agriculture trends.

In this paper we discussed about the problems and challenges related to Big Data analytics tools like Map Reduce over Hadoop and HDFS which helps farmers to better understand their crop status and the market trends and to take accurate decisions.

Hadoop:
Hadoop comprises of distributed file system, data storage and analytics flavours and a component that handles all computations, workflows, and configuration administration. HDFS runs among the nodes in a Hadoop cluster and together connects the file systems on many input and output data nodes to make them in to large big file system.
Here in this section we have given some of the tools and techniques to implement the big data analytics using agriculture data sets.

Technology	Description	In Agriculture
Hadoop	It is Open-source software framework for storing data and running applications on clusters of different	It is developed for large amount of agricultural data sets like crop images and video data in OLAP environment.

	hardware. It gives massive storage for any kind of data, efficient processing power and the ability to handle virtually limitless different tasks or jobs.	
Map Reduce	It is the Programming model and an associated implementation for analysing and creating large data sets with a parallel computation, distributed algorithm on different clusters.	An implementation of the Map Reduce programming model for large scale Agricultural data processing tasks.
HDFS	It is a Java-based file system that provides scalable and reliable data storage, and it was designed to span large clusters of commodity servers. HDFS has configured production scalability of up to 300 PB of storage and a single cluster of 4500 servers.	The core of Apache Hadoop consists of a storage part, known as Hadoop Distributed File System (HDFS), Stores large volumes of agricultural data like text, image and video data.
HIVE	It Supports analysis of large datasets stored in HDFS. The best part of HIVE is that it supports SQL-Like access to structured data which is known as HQL and big data analysis with the help of Map Reduce.	Hive provides a SQL-like interface to data stored in HDP. We can access the agricultural data depends upon the requirement. Hive provides a database query interface to Apache Hadoop.
HBase	It is designed for very large tables with large number of rows and columns; HBase is a distributed database that provides random real-time read or write access to big data.	HBase is a non-relational database and it contains large tables. So we can the store the data in large number of rows and columns

IV.IMPLEMENTATION

The implementation process consists of four modules. Figure shows the different modules.

1) Data Collection module
2) Data Storage module.
3) Analysis module.
4) Presentation module.
5) Decision Making Module

1) Data collection module:
This module collects data from various sources like sensor data, weather forecasting, text data and video data of a crop. These data can be issued manually, data can be acquired with access of data acquisition equipment, the received data are first stored in a database, when small data are collected, the small data will be transferred into the next module, transferred data will be automatically deleted.

2) Data storage module
The main responsibility of this module is for storage of metadata and data sets with replicated copy, which provide storage facility. HDFS is a storage container and is not restricted to any type of data. Small data in the data collection stage accumulated to a some amount will be placed in the storage phase on the actual basis.

3) Analytics module
This module contains two phases which includes data reading/analyzing phase and establishment of forecast result phase. The data reading is processed mainly by Hive. Hive is a framework for data processing on top of Hadoop. It was proposed to make it possible for data scientists with high SQL skills to run on the large volumes of data stored in HDFS. Hive runs on workstations and convert SQL queries into series of Map Reduce jobs for execution on a Hadoop cluster. Map Reduce is an execution engine suitable for huge data processing and can prominently improve the response speed for processing query results.

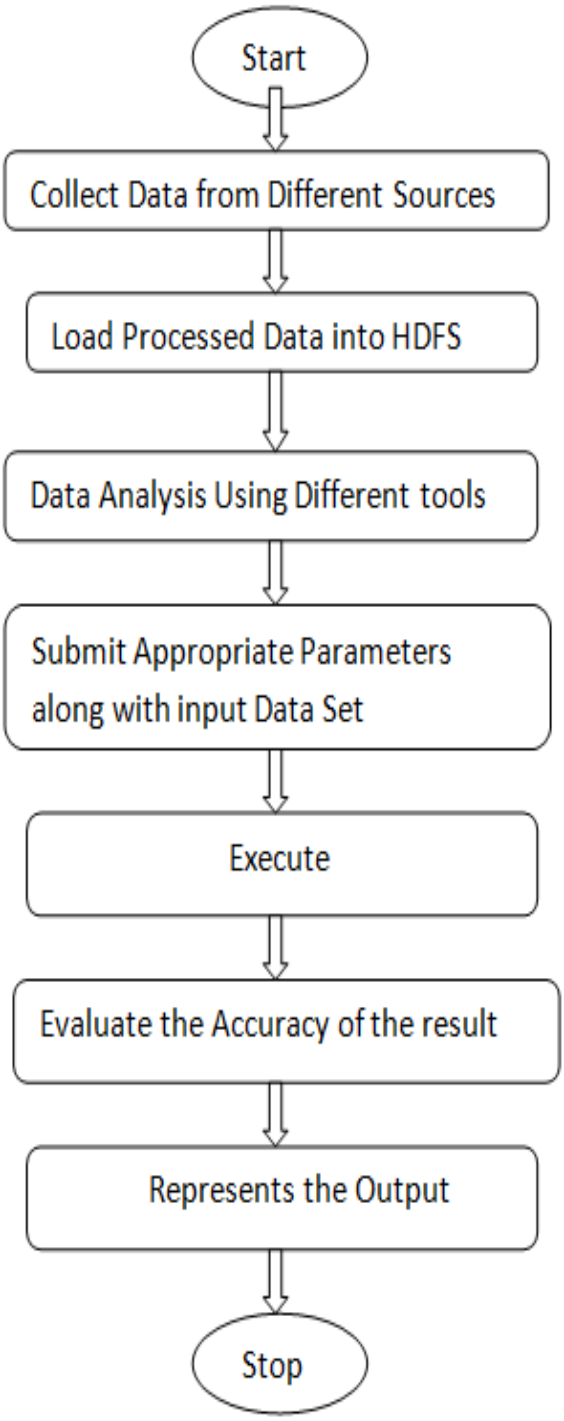
4) Presentation module

International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

The results which are returned from the query analysis module will be presented in this module in a visualized mode. The representation of complex data with charts and graphs is an essential part of the data analysis process, and we make use of different tools to create accurate and unique data visualizations.

Work flow of Proposed System

The work flow of the proposed system involves 7 steps. In each step, separate tasks are involved. The details of all the steps are given



In the first step datasets will be collected from various sources and these data sets are further pre-processed to make an effective input. After data cleaning, load into HDFS then apply query using Hive for analysis. We can also run Pig script for analyzing these data and result is giving as an input to visualizing tools to draw the graph for analysis. Then check the accuracy of the prediction.

Yield modelling

The yield of a crop can be estimated with following sources.

$Y=f(g, e, p) + \epsilon$ where

Y=yield
G=genetics
E=environment
P=practices
ϵ=variability

Description of Data Set:

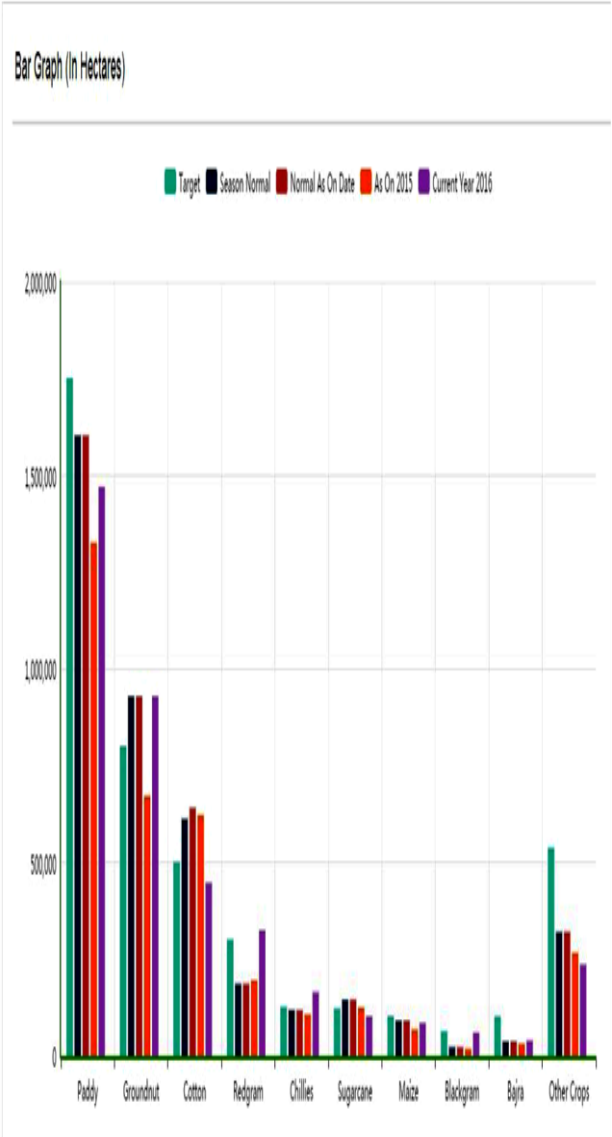
The given data set includes the information of variety of crops like Paddy, Jowar, Maize, Bajra, Horse gram, Raagi, Green gram etc. In this work we concentrate on the area of the crops sown, type of the crops and type output of the crop yield can be visualized in Andhra Pradesh state. The parameters considered are shown in the below table.

International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

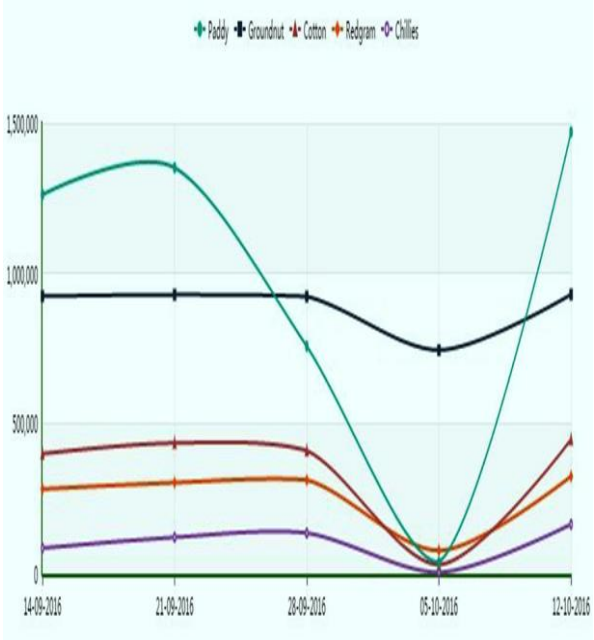
District	Crop	Normal (in Hectares)		As On 12-10-2015	Current Year	% Sown To	
		Season	As On Date	(in Hectares)	(in Hectares)	Season	As On Date
Anantapur	Paddy	26116	26116	7672	16471	63%	63%
Anantapur	Jowar	13102	13102	5278	1490	11%	11%
Anantapur	Bajra	2220	2220	2421	2410	109%	109%
Anantapur	Maize	21031	21031	11762	14010	67%	67%
Anantapur	Raagi	1513	1513	1149	1076	71%	71%
Anantapur	Minor Millets	2878	2878	8476	2465	86%	86%
Anantapur	Redgram	46864	46864	36343	67547	144%	144%
Anantapur	Greengram	6193	6193	12380	1535	25%	25%
Anantapur	Blackgram	187	187	25	1365	730%	730%
Anantapur	Horsegram	6415	6415	6280	812	13%	13%
Anantapur	Other Pulses	1418	1418	2062	986	70%	70%
Anantapur	Groundnut	629853	629853	442374	609377	97%	97%
Anantapur	Sesamum	0	0	0	0	0%	0%
Anantapur	Castor	17161	17161	9781	9308	54%	54%
Anantapur	Sunflower	5659	5659	1392	1029	18%	18%
Anantapur	Soyabean	1034	1034	158	616	60%	60%
Anantapur	Other Oil Seeds	1	1	0	0	0%	0%
Anantapur	Cotton	43622	43622	51953	30094	69%	69%
Anantapur	Mesta	6	6	0	0	0%	0%
Anantapur	Chillies	3182	3182	0	4248	134%	134%
Anantapur	Sugarcane	175	175	44	13	7%	7%
Anantapur	Onion	1995	1995	0	2265	114%	114%
Anantapur	Turmeric	90	90	0	0	0%	0%

V.RESULT ANALYSIS

This result analysis of overall crop harvested data set of Andhra Pradesh state over the past 2 years (2015 & 2016) to predict the data of the coming next year in advance, and uses the above crop prediction system which enables to be realized to make predictions. Figure below shows the graph representation of analysis. Here we use Pig script to analysis and the output is giving to as an input to generate a graph.



Here we analyze past 2 years crop harvested data of Andhra Pradesh state and result will be produced. In figure above, we analyze crops harvested information data sets and calculate year wise and monthly wise average crop harvesting for each variety of crop in each district of Andhra Pradesh region.



VI.CONCLUSION

Thus we have a tendency to conclude that, there's a growing range of applications and the role of Big data analytics techniques and internet of things (IoT) in agriculture and a growing quantity of information that area unit presently obtainable from several resources. This is often comparatively a completely unique analysis field and it's expected to grow within the future.

Future scope

In addition to the use of sensors and big data Analytics we can also use drones in the agriculture sector for better results. Drones can provide farmers with three types of structured views. First, seeing a harvesting crop from the air can display patterns that expose everything from irrigation issues to soil variation and even pest and fungal infections that are not seen at our eye level. Second, airborne cameras can take multispectral pictures, capturing data from the infrared and the visual spectrum of images, which can be combined to create a view of the harvested crop that highlights differences between healthy and infected plants in a way that can't be seen with the normal eye. Finally, a drone can suspect a crop every week, every day, or even every hour also. It also involves combined to create a time-series animation, that produced imagery can show changes in the crop, revealing trouble patterns or opportunities for better crop management.

REFERENCES

[1] 1.Environment Change Prediction to Adapt Climate-Smart Agriculture Using Big Data Analytics, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)

Volume 4 Issue 5, May 2015 Ramya M G ,Chetan Balaji ,Girish L

[2] 2..https://en.wikipedia.org/wiki/Agriculture_in_India

[3] [1] Prof. Chandrakanth. Biradar1, Chatura S Nigudgi," An Statistical Based Agriculture Data Analysis" International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 9, September 2012.

[4] [2] Farah Khan, Dr. Divakar Singh,"Association Rule Mining in the field of Agriculture: A Survey ", International Journal of Scientific and Research Publications, Volume 4, Issue 7, July 2014

[5] [3] Hetal Patel, Dharmendra Patel, "A Brief survey of Data Mining Techniques Applied to Agricultural Data"International Journal of Computer Applications,Volume 95– No. 9, June 2014.

[6] [4]P.Surya," the role of big data analytics in agriculture sector : a survey", International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST),Vol. 2, Special Issue 10, March 2016