2015 1st International Conference on Next Generation Computing Technologies (NGCT-2015)
Dehradun, India, 4-5 September 2015

Big Data in Precision Agriculture : Weather Forecasting for Future Farming

M. R. Bendre
Department of Information Technology
SGGSIE&T
Nanded, India
Email: bendremininath@sggs.ac.in

R. C. Thool
Department of Computer Sci. and Engg.
SGGSIE&T
Nanded, India
Email: rcthool@sggs.ac.in

V. R. Thool
Department of Instrumentation and Control
SGGSIE&T
Nanded, India
Email: vrthool@sggs.ac.in

***Abstract*—This paper gives an idea about how to discover additional insights from precision agriculture data through big data approach. We present a scenario for the use of Information and Communication Technology (ICT) services in agricultural big data environment to collect huge data. Big data analytics in agriculture applications provide a new insight to give advance weather decisions, improve yield productivity and avoid unnecessary cost related to harvesting, use of pesticide and fertilizers. Paper list out the different sources of big data in precision agriculture using ICT components and types of structured and unstructured data. Also discussed big data in precision agriculture, an ICT scenario for agricultural big data, platform, its future applications and challenges in precision agriculture. Finally, we have discussed results using a programming model and distributed algorithm for data processing and forecasting application of weather.**

Keywords*—Big data; big data analytics; precision agriculture; information and communication technology; weather forecasting.

I. Introduction

In precision agriculture, real time and historically generated data is collected in structured and unstructured datasets. As precision agriculture generates more data in the unstructured form and current research trend is to find knowledgeable information from them [1]. The need of future farming is to improve the quality of agricultural products and services by reducing investment cost. Big data can supports wide range of precision agriculture functions for discovering intelligence and insights from data to address many new and important farming decisions and problems. In the agriculture sector, ICT plays an important role to provide new technologies for data generation, transformation, and management [2]. New technology provides a framework for finding insights from data to give decisions regarding improve productivity and unwanted investment in advance.

Big data in agriculture refer to the huge amount of data, generated in agricultural practice and measurement. The processing and management of huge data is a challenging task over traditional methodologies and platforms. The processing of huge quantity of data needs a new hardware and software platform with tools and techniques. Mainly precision agriculture datasets having data related to crop patterns, crop rotations, weather parameters, environmental conditions, soil types, soil nutrients, Geographic Information System (GIS) data, Global Positioning System (GPS) data, farmer records, agriculture machinery data, such as yield monitoring and Variable Rate Fertilizers (VRT) [1][3][4]. The researchers working in this area have an opportunity to discover knowledge from huge data. The big data analytics can support to discover relationship, find patterns and trends from the data. The big data can solve various agricultural applications like decision making, extract new insights and knowledge for future farming. The ICT provides information to farmers through mobile apps, Short Message Service (SMS) services, agriculture knowledge hubs and new generation web applications. The ICT provides research equipment's to the researchers for the precision agriculture, remote sensing such as GPS, GIS, devices and data monitors. Also, ICT plays a key role in the remote data transmission and management.

The rest of this paper is organized as follows. Section II provides background information, including types of data, characteristics of big data in precision agriculture. Section III to VI provides agricultural big data collection scenario, applications, management and challenges in precision agriculture. Section VII and VIII synthesizes methodology for identification of patterns and processing of big data in precision agriculture for weather data and the use of this methodology is illustrated using a case study. The last section IX brings main conclusions, and outlines possible directions for future work.

II. Big Data in Precision Agriculture

The maximum data in the agriculture sector are generated by the on-site farming, remote farming or satellite farming called as precision agriculture. The world population increasing day by day and it is expected to reach 10 billion in next 35 years. To feed the world population, essential development in the agricultural production and disaster management are now important [5]. It is essential to gather data from various sources of precision agriculture for predictive decisions. The gain in agricultural production and management are depend on weather conditions. So prediction of an area based weather is an essential task for future farming in precision agriculture.

A. Types of Data

Agricultural big data are collected in the form of structured and unstructured type. The homogeneous and heterogeneous sensing devices and new technologies are essential to gather such diverse type of data. Typically precision agriculture dataset refers to the following types:

1) Historical Data: It includes, soil testing, crop patterns, field monitoring, yield monitoring, climate conditions, weather conditions, GIS data, and labor data.

2) Agricultural Equipment's and Sensor Data: It includes data collected from remote sensing devices, GPS based receivers, variable rate fertilizers, soil moisture's, temperature sensors, farmers call records and equipment logs.

3) Social and Web Based Data: It includes, farmers and customers feedback, agricultural websites and blogs, social media groups, web pages, and data from search engines.

4) Publications: It includes, agriculture research and agriculture reference material such as text-based practice guidelines and agriculture requirements (e.g. pesticides, fertilizers, and equipment's information) data.

5) Streamed Data: It includes data from, crop monitoring, mapping, drones, aircraft's, wireless sensors, smart phones, security surveillance's.

6) Business, Industries and External Data: The data from billing and scheduling systems, agriculture departments and other agriculture equipment manufacturing companies.

B. Characteristics of Big Data

The main three characteristics of the big data are as follows:

1) Volume: The amount of data generated by organizations or individuals is in the big size such as petabytes to zettabytes. Enterprises in all industries are looking for the ways to handle the ever increasing volume of data that's being created every day.

2) Velocity: The frequency and speed at which data is generated, captured and shared. Farmers as well as business related to agriculture are now generating more data and in much shorter cycles, from hours, minutes, seconds down to milliseconds.

3) Variety: The proliferation of new data types including those from social, machine and mobile sources. New types include content, location or geo-spatial, log data, maps, machine data, sensor data, mobile, physical data points, process, Radio Frequency Identification (RFID), search, sentiment, streaming data. The variety includes traditional unstructured data also.

III. Scenario of Big Data in Precision Agriculture

Precision agriculture through ICT consist of different layers such as application layer, store and processing layer and infrastructure layer. In the application layer, data acquisition tools, web based solutions, software's and development platforms are present. The storage and management of big data need a novel system and platform, today's cloud computing solutions provide such huge amount of storage and management [6]. The distributed and parallel systems make a role in the data processing and management. The MapReduce based model was used for the big data processing. Mainly ICT plays a role of data acquisition, management and visualization over the world wide in different applications [7]. The overall scenario and components in the precision agriculture shown in the Fig. 1. At last, infrastructure layer consists of clustered network of sensors and systems which are used to generate, access and manage large amount of data.

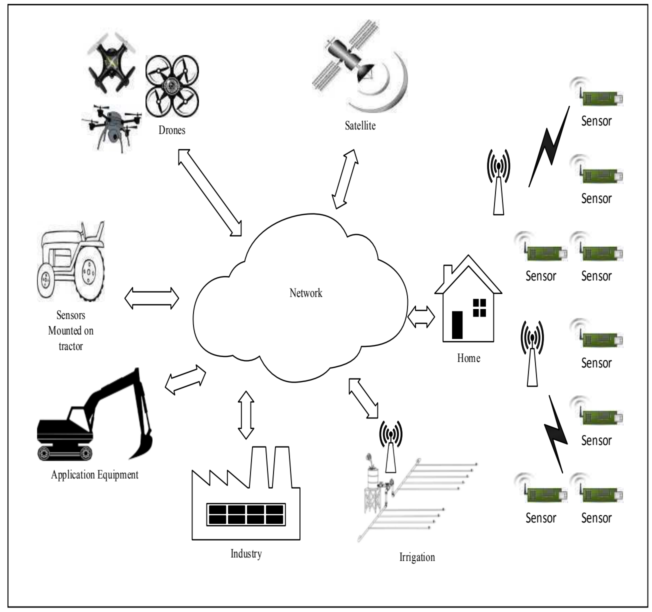

Fig. 1. ICT scenario for precision agriculture.

A. Big Data Analytics Platform

Hadoop [8] developed by Apache is a parallel and distributed platform over the ICT using a large number of the commodity hardware's for the big data processing. The components of Hadoop help for data acquisition, management and visualization of applications in the ICT. Mainly, Mahout-machine learning applications, a MapReduce-distributed and parallel processing, Hbase-structured and unstructured data storage and management, Drill and Strom-interactive analysis and steam processing [8][9]. Hadoop Distributed File System (HDFS) is used for storing large amount of data. The MapReduce programming model is used for parallel and distributed processing of data.

IV. ICT applications in Precision Agriculture

In the precision farming ICT plays an important role using technological tools such as hardware, software and practices to gather data, process, visualize and decision making. There are the various practices used for the data collection and processing to make decisions in the precision agriculture.

A. Global Positioning System (GPS)

GPS satellites calculate the position on the earth using receivers. These devices are used for the real-time data acquisition at the time of soil and crop measurement. These devices are mounted in the field with various agriculture equipment's to take signal for measurement. Communication technology for measuring data on field increases the accuracy using GPS signals.

B. Remote Sensing (RS)

The sensors play a role in precision agriculture for collection of huge and variety of data from remote locations. The wireless sensor network is used for the data acquisition for e.g. sensing soil moisture and temperature [10]. The wired and wireless sensors mounted on agriculture equipment's, drones, aircraft's, or satellite and grounded in the field are used for plant monitoring, soil sensing, disaster management and many agricultural applications. They can check plant stress, moisture, humidity, nutrients, disease and other crop health parameters by analyzing images and data collected from the remote sensing devices. Currently different companies enter in the precision agriculture sector to provide services using their own sensing devices.

C. Geographic Information System (GIS)

GIS is used to store yield maps, sensed data soil survey reports, soil nutrient levels. GIS is the collection of software and hardware modules for generation of the maps. These maps can help with various agriculture management strategies.

D. Variable Rate Fertilizer (VRT)

To increase the soil fertility soil sampling is recommended, by taking different soil samples and as per soil nutrients make graphical visualization that decides the use of fertilizers. These maps with GPS signal are input to the fertilizer spreader for future management. Best use of ICT in agriculture is to increase production and avoid misuse of fertilizers.

V. Precision Agriculture Management

ICT approach in precision agriculture management gives a novel technology and platform to farmers, government departments, industries and researchers. Following are the management strategies used in the precision agriculture [4][11][12]. There are different big data functionality used for various management applications in precision agriculture.

A. Farming Decision Support

Big data analytics and ICT technology supports to acquire, understand, categorize and discover information from large amounts of data. Also, it can predict future or recommended decisions to farmers and vendors at the point of precision agriculture. It can help farmers to manage and increase yield productivity.

B. Water Management

Predictive data mining or analytic solutions over ICT can leverage water management and automatic irrigation (e.g., as per soil humidity and new technology of irrigation) in real time to improve best practices to crops.

C. Increase Productivity

Web and mobile based applications visualize information from historical data, crop patterns and weather data. Big data analytics and ICT solutions can also support agriculture equipment companies and departments to perform analysis over agricultural growth and productivity, to support and identify future farming trends.

D. Agriculture Disaster Management

Big data analytics and ICT applications can take initiatives such as real time management in precision agriculture, where it can mine knowledge from historical unstructured data, discover patterns to predict events that are harmful in farming. These patterns and decisions may help farmers in the disaster management in agriculture.

E. Policy, Financial and Administrative

The analysis supports policy makers, service providers, companies and government departments for deciding future varieties, pesticides and fertilizers.

VI. Big Data and ICT Challenges in Precision Agriculture

The main challenges are discovering knowledge and correlations from historical records, understanding big data, unstructured data in the proper format, handling huge amount of imaginary and video data, handling data of crop monitoring through several sensors and their various interactions and communications, adoption and accessibility of new generation technologies for the individual farmer is an expensive task. In the agriculture community lack of low technology knowledge need more training, security and management of ICT equipment's [13][14].

VII. Methodology and case Study

The companies are investing millions of dollars in precision agriculture to increase yield production. So weather forecasting for better agriculture decision and production in the green zone or dry area is essential. In this case study daily minimum, maximum temperature, humidity and rainfall data of the KVR (Krishi Vidyapeeth Rahuri (KVR), Ahmednagar, India) weather station from last 10 year, were collected and analyzed. All the parameters used in the case study are rainfall in millimeter (mm), temperature in degree Celsius (^oC) and humidity in percentage (%). KVR station is located at 19.38^oN and 74.65^oE. It is 3 km and 15 km far from the two main rivers of the Ahmednagar district Pravara and Mula respectively. Huge rainfall in the area of two dams namely Bhandardara and Muladam, so these rivers overflow each year. This area of agriculture is in the green zone and the half of year these rivers flow with water. So this weather based prediction and decision support study for the agriculture in this area is important. The study data consist of daily weather parameters collected from KVR rain gauge between 1 Jan 2003 to 31 Dec 2013.

A. Working with Big Data

To increase the capacity of memory and execution speed 64-bit computing is better than 32 bit computing. This allows to access big data from disk very fast. Storage functions and capabilities used to import and access data from files, tables and set of files. By running parallel processor multiple cores of computer give better performance on big size of data. Parallel computing algorithms run on different cores that can scale up computing. Amazons Elastic Computing Cloud (EC2) is solution for Cloud Computing in parallel and distributed processing of the big data over thousands of machines [15].

The MapReduce programming technique is used for analysis of data that not fit in the memory. Machine learning algorithms and methods can be used for the development of the predictive model on big amount of data. The decision trees, K-means, clustering, K-nearest neighbor search, the expectation maximization algorithm, hidden Markov models, and neural networks are the mainly used machine learning algorithms [16]. For predictions time series methods are the best solutions. Supervised and unsupervised machine learning methods with big data technology can be used for the discovering insights from the large data set. The proposed predictive model consist of the MapReduce for bid data processing and linear regression for data prediction. Hadoop with the MapReduce functionality support to access data using data retrieval methods and apply this data to the MapReduce programming environment for further processing [17]. The MapReduce programming model is used to solve the big data problem by distributing it over the distributed file system. Google File System (GFS) is used to distribute data over the clustered network or distributed platform for the processing. The MapReduce support to directly manage distribution and processing of data [18][19].

B. Working with Predictive Analytics

Predictive analytics are used to extract knowledge from the existing data used for future planning and also gives idea about trends and outcomes. Predictive analytics is the practice of extracting information from existing data sets in order to determine patterns and predict future outcomes and trends. Traditional weather forecasting has been built on a foundation of deterministic modeling start with initial conditions, put them into a supercomputer model, and end up with a prediction about future weather [20]. The approach, ensemble forecasting was introduced in the early 1990s in which the number of systems is used to process data. In concert with statistical techniques, ensembles can provide accurate statements about the uncertainty in daily and seasonal forecasting. The challenge now is to improve the modeling, statistical analysis, and visualization technologies for disseminating the ensemble results.

C. Linear Regression

The regression models are used in the study to handle large linear and nonlinear data. It is the method of supervised learning for prediction of the future patterns from historical data items. Where by calling the numerical dependent variable y, and finds approximate value of the numerical values x. The data collected from the KVR station are stored in the database by using data storage functions and can access number of data vectors for processing. Where the vector x consisting of n numerical values $(x_1, x_2, x_3..., x_{n-1}, x_n)$, and n is number of feature values of each data item in the dataset. The equation 1 shows general model used for the data processing.

$$\widehat{y} = f(x) + \xi \tag{1}$$

The equation 1 gives the difference between the actual and predicted value is denoted by ξ. The predicted value of y is $f(x)$ and is indicated by $\widehat{y}$ symbolically. As linear regression could be used when there is a linear (or roughly linear) dependency between x and y. In this case equation 2 shows the algorithm used to model y as a function of x:

$$\widehat{y_i} = a_0 + a_1 x_i + \xi \tag{2}$$

To calculate the regression coefficients a_0 and a_1 the equation 4 and 5 are used.

$$\bar{xy} - \bar{x}\bar{y} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \tag{3}$$

$$a_1 = \frac{\bar{xy} - \bar{x}\bar{y}}{y^2 - \bar{y}^2} \tag{4}$$

$$a_0 = \bar{y} - a_1\bar{x} \tag{5}$$

The size of the x and a_1 data vectors are equal to the number of dimensions in the feature space and error term ξ is the difference between the actual and predicted value of the target variable. The y symbol is use to indicate the predicted value of target variable by the model and we have $y = a_0 + a_1 x$. The algorithms to learn a_0 and a_1 values of the items in the dataset. The objective is to minimize the difference of actual and predicted values for all data items (as an example, this difference could be measured by minimizing the sum of square of the difference between actual and predicted target values):

$$f(q) = \frac{1}{N}\sum_{i=1}^{N}\xi_i^2 \tag{6}$$

D. MapReduce Algorithm

The aim of MapReduce algorithm is to split the chunk of data for faster processing. The MapReduce is used to calculate mean values and that are used in the regression model for fitting and calculating the trends of data that is predicted values. The KeyVS, VIterator, datastore, hasnext, getnext, VariableNames, iKVStore, oKVStore are the variables used in the MaprReduce programming in this section. MapReduce moves each chunk of data in the input datastore through several phases before reaching the final output. The Fig. 2 shows outline, the phases of the algorithm for MapReduce. The MapReduce algorithm reads a chunk of data from the input datastore using read object method and then calls the map function to work on that chunk. The map function receives the chunk of data, organizes it or performs a precursory calculation, and then uses the functions to add key-value pairs to an intermediate data storage object called a KeyVS. The number of calls to the map function by MapReduce is equal to the number of chunks in the input datastore. After the map function works on all of the chunks of data in the datastore, MapReduce groups all of the values in the intermediate KeyVS object by a unique key. Next, MapReduce calls the reduce function once for each unique key added by the map function. Each unique key can have many associated values. MapReduce passes the values to the reduce function as a VIterator object, which is an object used to iterate over the values. The VIterator object for each unique key contains all the associated values for that key.

The reduce function uses the hasnext and getnext functions to iterate through the values in the VIterator object one at a time. Then aggregating the intermediate results from the map function, the reduce function adds final key-value pairs to the output using functions. The order of the keys in the output is the same as the order in which the reduce function adds them to the final KeyVS object. MapReduce does not

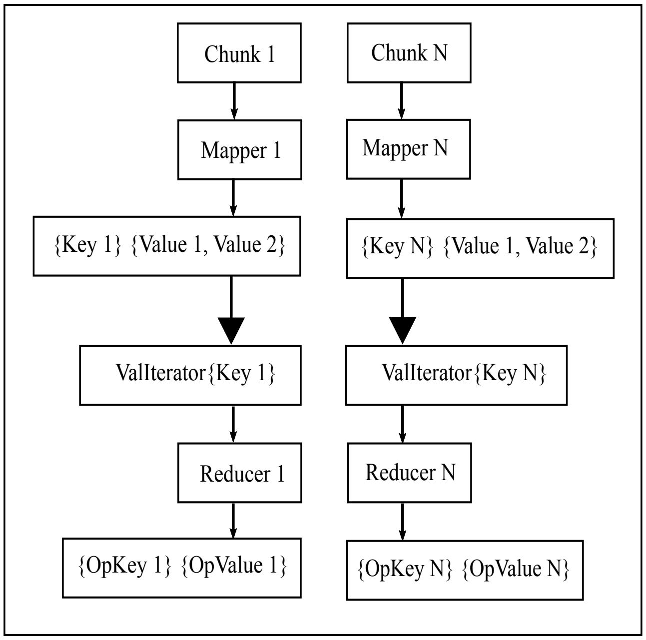

Fig. 2. MapReduce processing.

explicitly sort the output. The MapReduce function executes read on the datastore to retrieve data to pass to the map function. Therefore, VariableNames, Formats, and ReadSize options to directly configure the chunk size and type of data that MapReduce passes to the map function. To select the temperature and rainfall variable as the only variable of interest, specify VariableNames following object is used.

datastore.VariableNames = ‘temperature’;

Now, whenever the read, readall or preview functions act on datastore, they will return only information for the rainfall variable. This allows to examine the format of the data and then MapReduce function will pass to the map function. The MapReduce function automatically calls the map and reduce functions during execution, so these functions must meet certain requirements to run properly. The inputs to the map function are data, info, and iKVStore: data and info are the result of a call to the read function on the input datastore, which MapReduce executes automatically before each call to the map function. The iKVStore is name of intermediate KeyVS object to which the map function needs to add key-value pairs. The add and addmulti functions use this object name to add key-value pairs. If none of the calls to the map function add key-value pairs to iKVStore, then MapReduce does not call the reduce function and the resulting datastore is empty. A map function is:

```
function MeanTempMapFun(data, info, iKVStore)
temp = data.temperature( isnan(data.temperature));
sumLenValue = [sum(temperature) length(temp)];
add(iKVStore, ‘sumAndLength’, sumLenValue);
end
```

This map function perform some straightforward roles. The first line filters out all NaN values in the chunk of distance data. The second line creates a two-element vector with the total distance and count for the chunk, and the third line adds that vector of values to iKVStore with the key, ‘sumAndLength’. After this map function runs on all of the chunks of data in datastore, the iKVStore object contains the total distance and count for each chunk of distance data. The inputs to the reduce function are iKey, iValIter, and oKVStore: iKey is for the active key added by the map function. Each call to the reduce function by MapReduce specifies a new unique key from the keys in the intermediate KeyVS object. iValIter is the VIterator associated with the active key, iKey. This VIterator object contains all of the values associated with the active key. Scroll through the values using the hasnext and getnext functions. oKVStore is the name for the final KeyVS object to which the reduce function needs to add key-value pairs. MapReduce takes the output key-value pairs from oKVStore and returns them in the output datastore, which is a KeyValueDatastore object by default. If none of the calls to the reduce function add key-value pairs to oKVStore, then MapReduce returns an empty datastore. A reduce function is:

```
function MeanTempReduceFun(iKey, iValIter, oKVStore)
sumLen = [0 0];
while hasnext(iValIter)
sumLen = sumLen + getnext(iValIter);
end
add(oKVStore, ‘Mean’, sumLen(1)/sumLen(2));
end
```

This reduce function loops through each of the rainfall and count values in iValIter, keeping a running total of the distance and count after each pass. After this loop, the reduce function calculates the overall mean rainfall with a simple division, and then adds a single key to oKVStore. The calculation of the average rainfall and temperature from the data set, call MapReduce using datastore. Logistic regression is a way to model the probability of an event as a function of another variable. In this example, logistic regression models the probability of a maximum temperature being more than normal. To accomplish this regression, the map and reduce functions must collectively perform a weighted least-squares regression based on the current coefficient values. The mapper computes a weighted sum of squares and cross product for each chunk of input data.

VIII. Results and Discussion

This paper presents a scenario for big data and methodology for forecasting of weather data. The objective of this paper is to increase the accuracy of the forecasting using different weather parameters for the future precision agriculture. The scenario can be used to gather big data using various ICT components. The preprocessing techniques make superior data for forecasting model. This model shows the best performance with different datasets and the purpose of this study was to investigate the effective model coupled with data preprocessing techniques to improving the accuracy of rainfall forecasting. The methodology is illustrated using a case study for weather forecasting data collected between January 2003 and December 2013 at a weather station located in a green zone region of Ahmednagar. The historical data applied to the processing model in this study, that is temperature and rainfall through MapReduce, to find out mean and average values. To process big data MepReduce is used to minimize time of execution and produce as early as results for decision making in precision agriculture. The results are calculated using the model by applying different rainfall values, when the historical data values increases, it minimizes the execution time as per normal execution. The model predicts rainfall and temperature values for the year 2013 and also compared actual and predicted values to minimize the error. Table 1 shows the temperature T in Celsius and rainfall R in millimeter compared using predicted values used for season identification.

Fig. 3 and 4 shows the input data of historical rainfall and temperature data used for the prediction study and different color denotes the year wise data from 1^{st} January to $31^{s}t$ December.

TABLE I. WEATHER CONDITIONS

Months	Temperature	Rainfall	Season
July-Oct	$25 > T < 35$	$10 > R < 50$	Monsoon
Nov-Feb	$20 > T < 35$	$0 > R < 10$	Winter
Mar-June	$35 > T < 45$	$0 > R < 10$	Summer

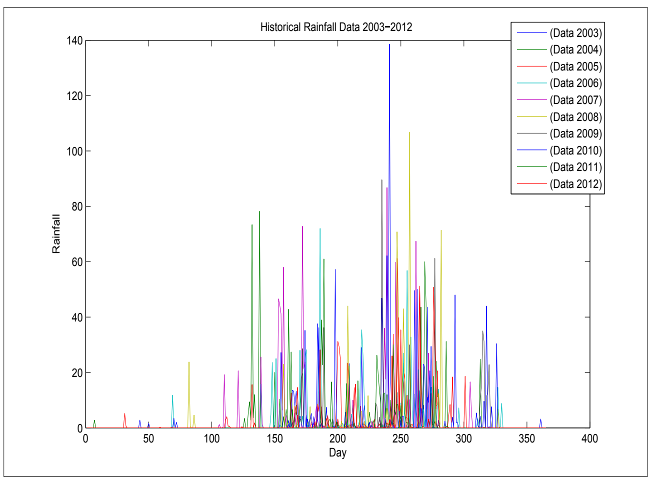

Fig. 3. Input historical rainfall data.

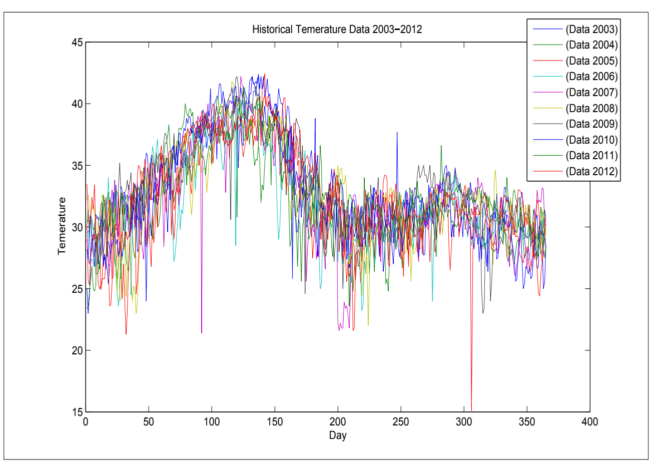

Fig. 4. Input historical temperature data.

Fig. 5 and 6 shows the actual data and predicted data of rainfall and temperature for the year 2013. Fig. 7 shows the regression R is calculated from the target and output values. Fig. 8 shows the probability of maximum temperature using MapReduce functions.

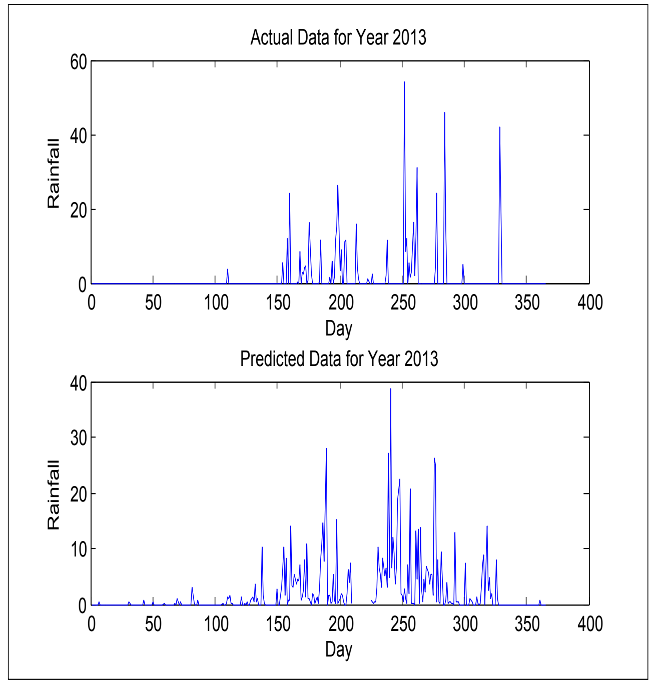

Fig. 5. Actual and predicted rainfall data for year 2013.

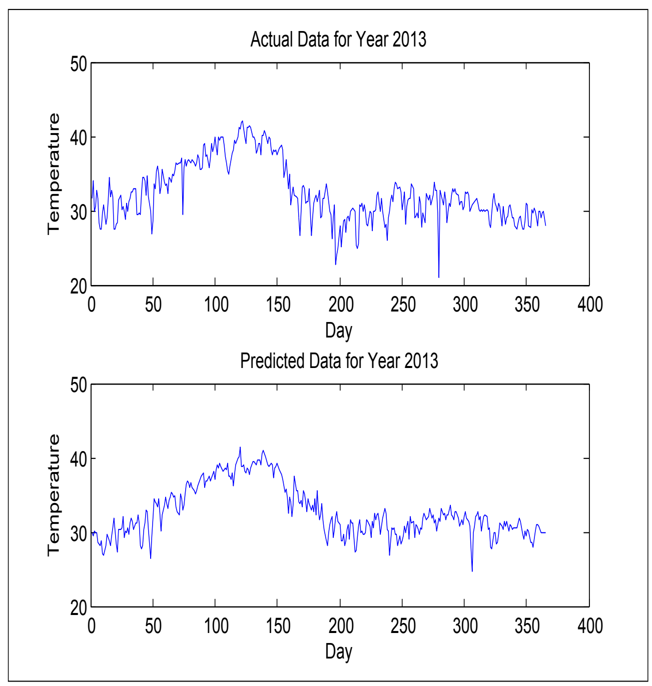

Fig. 6. Actual and predicted temperature data for year 2013.

IX. CONCLUSION

Big data analytics and ICT in agriculture are evolving technologies into a promising field for providing insight from very large data sets and improving productivity and reducing

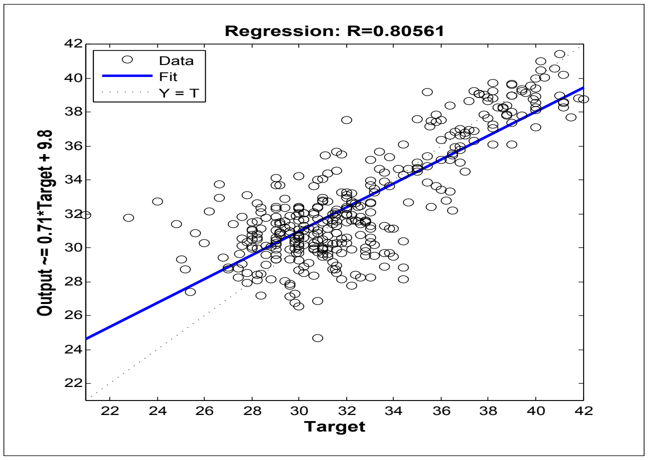

Fig. 7. Regression output.

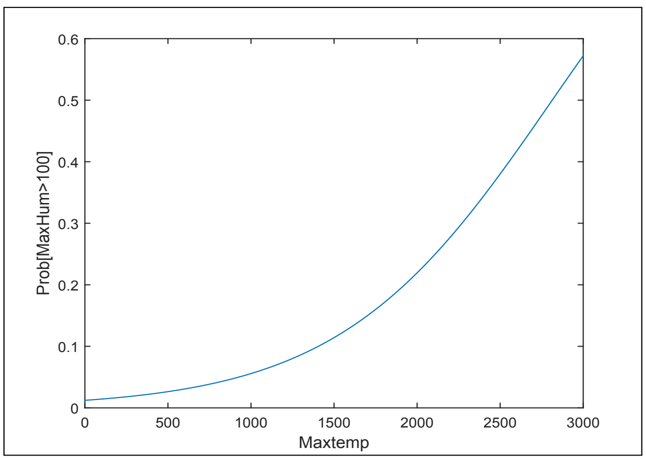

Fig. 8. Probability of maximum temperature with humidity.

investment costs. Big data analytics and ICT has the potential to uses novel technologies and platform to generate, collect, process and visualize large data for future predictions and make decisions. In the precision agriculture, remote sensing devices plays a vital role in data collection and real time decision support. The results forecast using a regression model and big data handle by Mapreduce of this study shows a considerable potential of data fusion in field of crop and water management for applications such as precision agriculture. As per these results model predict the temperature and rainfall in the region of case study. It suggests various decisions to farmers for deciding the crop pattern and water management in the future. It is solution for the yield management and disaster management to increase the gain of food production. In the future we'll see the rapid growth and use of big data analytics through ICT across the agriculture organization and the agriculture industry to increase yield production. Big data analytics and ICT applications in precision agriculture are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process to increase productivity of agriculture.

ACKNOWLEDGMENT

The authors would like to thank Dr. C. D. Deokar & Dr. P. B. Kharde from Krishi Vidyapeeth Rahuri, and Dr. B. Gaikwad form Krishi Vigyan Kendra Babhaleshwer for agricultural guidance and data support for this research.

REFERENCES

[1] R. D. Grisso, M. M. Alley, P. McClellan, D. E. Brann, and S. J. Donohue, "Precision farming. a comprehensive approach," 2009.

[2] R. D. Ludena, A. Ahrary *et al.*, "Big data approach in an ict agriculture project," in *Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on*. IEEE, 2013, pp. 261–265.

[3] B. Brisco, R. Brown, T. Hirose, H. McNairn, and K. Staenz, "Precision agriculture and the role of remote sensing: a review," *Canadian Journal of Remote Sensing*, vol. 24, no. 3, pp. 315–327, 1998.

[4] S. W. Searcy, *Precision farming: A new approach to crop management*. Texas Agricultural Extension Service, Texas A & M University System, 1997.

[5] K. Cukier and V. Mayer-Schoenberger, "Rise of big data: How it's changing the way we think about the world, the," *Foreign Aff.*, vol. 92, p. 28, 2013.

[6] B. Rao, P. Saluia, N. Sharma, A. Mittal, and S. Sharma, "Cloud computing for internet of things & sensing based applications," in *Sensing Technology (ICST), 2012 Sixth International Conference on*. IEEE, 2012, pp. 374–380.

[7] F. Awuor, K. Kimeli, K. Rabah, and D. Rambim, "Ict solution architecture for agriculture," in *IST-Africa Conference and Exhibition (IST-Africa), 2013*. IEEE, 2013, pp. 1–7.

[8] A. B. Patel, M. Birla, and U. Nair, "Addressing big data problem using hadoop and map reduce," in *Engineering (NUiCONE), 2012 Nirma University International Conference on*. IEEE, 2012, pp. 1–5.

[9] D. Borthakur, "The hadoop distributed file system: Architecture and design," *Hadoop Project Website*, vol. 11, no. 2007, p. 21, 2007.

[10] S. Nandurkar, V. Thool, and R. Thool, "Design and development of precision agriculture system using wireless sensor network," in *Automation, Control, Energy and Systems (ACES), 2014 First International Conference on*. IEEE, 2014, pp. 1–6.

[11] B. Venkatalakshmi and P. Devi, "Decision support system for precision agriculture," *International Journal of Research in Engineering and Technology*, vol. 3, no. 7, pp. 849–852, 2014.

[12] R. Saravanan, *ICTs for Agricultural Extension: Global Experiments, Innovations and Experiences*. New India Publishing, 2010.

[13] R. Khosla, "Precision agriculture: challenges and opportunities in a flat world," in *19th World Congress of Soil Science*, 2010.

[14] A. McBratney, B. Whelan, T. Ancev, and J. Bouma, "Future directions of precision agriculture," *Precision Agriculture*, vol. 6, no. 1, pp. 7–23, 2005.

[15] L. Wang, G. Von Laszewski, A. Younge, X. He, M. Kunze, J. Tao, and C. Fu, "Cloud computing: a perspective study," *New Generation Computing*, vol. 28, no. 2, pp. 137–146, 2010.

[16] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[17] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 1, pp. 97–107, 2014.

[18] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The google file system," in *ACM SIGOPS operating systems review*, vol. 37, no. 5. ACM, 2003, pp. 29–43.

[19] "How to work with huge and fast data sets," http://http://in.mathworks.com/discovery/big-data-matlab.html, 2015.

[20] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *Power Systems, IEEE Transactions on*, vol. 17, no. 3, pp. 626–632, 2002.