Environment Change Prediction to Adapt Climate-Smart Agriculture Using Big Data Analytics

Ramya M G
PG Student
Dept of CSE, Cit, Gubbi,
Tumkur, India

Chetan Balaji
Assistant Professor,
Dept of CSE, Cit, Gubbi,
Tumkur, India

Girish L
Assistant Professor
Dept of CSE, Cit Gubbi,
Tumkur, India

Abstract **—Now a day, the increased climate variability is challenging to farmers. It will expect to influence on crop and livestock productivity. So new approaches are required to farmers with updated and relevant information so as to support them in the decision making process, and to make them more resilient to climate change. The main objective of this project is to develop an application that predicts the environmental changes and to assist farmers in adapting to climate-smart Agriculture systems using big data Approach which increase the income and productivity of farmers. This paper is aiming to find Climate-Smart, Site-Specific Agriculture Decision-Making to problems, such as climate change-induced global food insecurity, to predicting and mitigating the impact of extreme weather events on global finance and also we can predict the Economic Dimensions of Climate Change around the world using Data and Analytics.**

This paper presents an enhanced automated prediction technique based on hadoop framework for efficient and scalable weather data analysing and forecasting system.

Index Terms- **prediction, climate-smart agriculture, decision making process, big data analytics, Hadoop framework**

I. INTRODUCTION

Agriculture is the backbone of economic system of a country. The evolution of an agriculture sector contributes to marketable surplus. Nation's export trade depends largely on agricultural zone. Stable agriculture system ensures a global food security. Agriculture systems are dynamic and managed ecosystems. Climate is the primary contributes of agriculture. Climate variability is expected to influence on crop and livestock production. Human response is critical to understanding and estimating the effects of climate change on production. The climate change like higher temperature has been found to reduce yields and quality of many crops. So new approaches are required for farmers with updated and relevant information so as to support them in the decision making process, and to make them more resilient to climate change.

We proposed an automatic prediction model which analyzes the large data sets of historical data using big data analytics. Big data analytics is the process of examining large amount of data comes from variety of sources like sensors data, weather forecasting, and social media data with variety of formats to uncover the hidden patterns, unknown correlations and also useful valuable information [1].

Economic models also typically estimate changes in market trends and conditions under climate variation. The project is aiming to find solutions to problems, such as climate change-induced global food insecurity, to predicting and mitigating the impact of extreme weather events on global finance. Since we have sensors everywhere in space which are used to monitor and measure weather, plant life, cloud cover, ice cover, precipitation, sea surface temperature, and many more geophysical parameters. These wide-ranging data collections give us increasingly deeper and broader coverage of climate change, both temporally and geospatially.

The remainder of this paper is as follows: In section 2, we offer related work on weather prediction. In section 3, we describe proposed system architecture with its applications. The implementation of prediction algorithm is shown in section 4 and finally Section 5 concludes overall article.

II. RELATED WORK

The statistical forecasting models can be developed to estimate the future weather conditions. The model is simple due to the fact that it uses simplemathematical equation using Multiple Linear Regression (MLR) to predict maximum temperature, minimum temperature and relative humidity [2]. Machine learning techniques achieved better performance than traditional statistical methods. So present Support Vector Machines (SVM) for weather prediction [3] and enhanced support vector regression model with more un-interpretable kernel functions in the domain of forecasting the weather conditions [4].

IoT based agriculture production system monitoring system to analyze crop environment and also provides method to improve the efficiency of decision making by analyzing harvest statistics, forecast agriculture production using IoT sensors [5]. Prediction model based on macro climate [6] relies on historical data and an analytical algorithm with field monitoring and traceability model.

Local time series analysis innovate statistical model and NWP introduces dynamical model of weather processes, another approach combines these two methods and introducing hybrids model. NWP based accurate localized short term weather prediction system [7] predicting the values of meteorological variable. Another Prediction model based on big data analysis implements hybrid technique which innovates a FCM clustering algorithm for complex characteristics of the industries [8].

ISSN: 2278 – 1323

III. PROPOSED SYSTEM

The proposed model is to find solutions to problems, such as climate change-induced global food insecurity, to predicting and mitigating the impact of extreme weather events on global finance. Since we have sensors everywhere in space which are used to monitor and measure weather, plant life, cloud cover, ice cover, precipitation, sea surface temperature, and many more geophysical parameters. These wide-ranging data collections give us increasingly deeper and broader coverage of climate change, both temporally and geospatially.

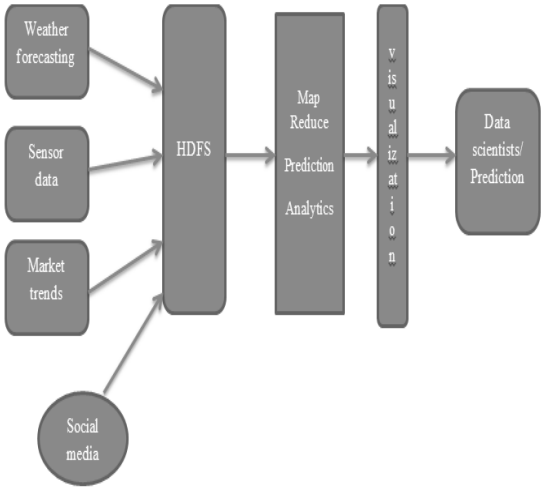

Figure 1: Architecture of proposed system

This impressive sensor delivers a vast rate of unstructured data (petabytes of data from daily). These massive unstructured data require the typical big data tools for data storage, processing, analyzing, visualizing, and predicting [9].

Figure 1 shows the architecture of the proposed system in which MapReduce in the use of Hadoop to monitor and analyzing the big data collected from various sources like weather forecasting, sensor data, market trends and social media data, then result has to be presented to Data scientist for predict future weather with high-resolution simulations, to assist farmers in adapting to climate change, and to protect the world's agriculture business.

IV. IMPLEMENTATION

4.1 The weather Forecasting Platform

The implementation process consists of five modules. Figure 2 shows the different modules.

- Data Acquisition module.
- Data Storage module.
- Query Analysis module.
- Presentation module.

A. Data Acquisition Module

This module collects data from various sources like sensor data, weather forecasting, social media data and market trends. These Meteorological data can be issued manually, data can be acquired with access of meteorological data acquisition equipment, the small received data are first stored in an oracle database, when small data are gathered to a certain number, the small data will be transferred into the storage module, transferred data will be automatically deleted.

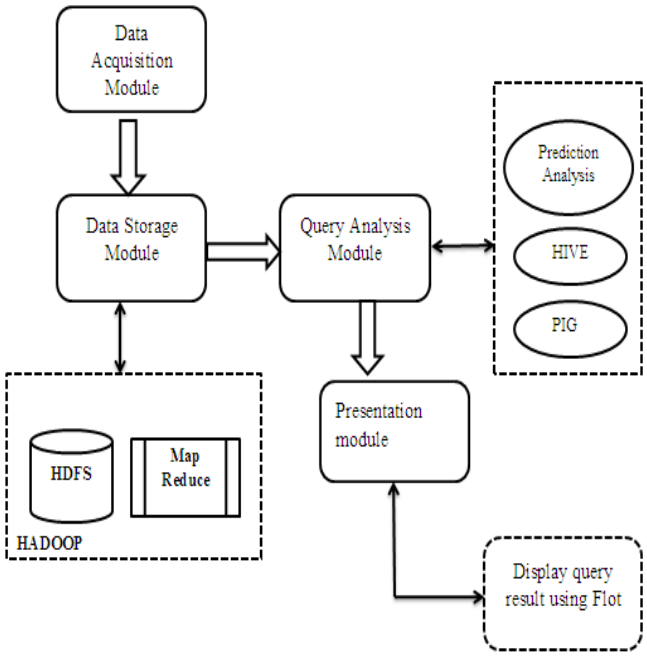

Figure 2: proposed system framework

B. Data storage module

It is responsible for storage of metadata and data sets with replicated copy, which provide backup facility. HDFS is a storage container and is not limited by any type of data. Small data in the data acquisition module accumulated to a certain amount will be placed in the storage module on a regular basis.

C. Query Analysis module

This module is processing phase includes two parts of data reading/analyzing and establishment of forecast result. The data reading is done mainly by Hive. Hive is a framework for data warehousing on top of Hadoop. It was created to make it possible for analysts with strong SQL skills to run on the huge volumes of data stored in HDFS. Hive runs on workstations and convert SQL queries into series of MapReduce jobs for execution on a Hadoop cluster. MapReduce is an execution engine suitable for large data processing and can significantly improve the response speed for returning query results. In the second part we implement prediction function for establish forecast data through k-means cluster algorithm. Here we make use of apache mahout. It is an open-source scalable machine learning library. Mahout provides an efficient way of implementing unsupervised machine learning algorithm. The data of past few years are used to make prediction about future.

D. Presentation module

The results which are returned from the query analysis module will be displayed in this module in a visualized mode. Representing complex data with charts and graphs is an essential part of the data analysis

ISSN: 2278 – 1323

process, and we make use of flotend tool to create beautiful and unique data visualizations.

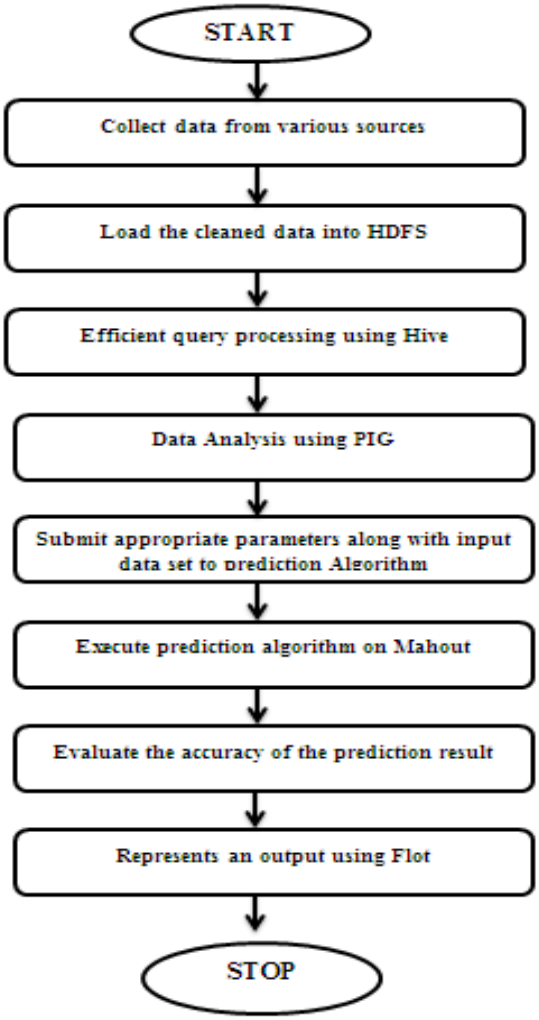

Figure 3: work flow of proposed system

Proposed system workflow is shown in figure 3. In the first step weather datasets will be collected from various sources and these data items are further preprocessed to make an effective input to prediction algorithm. 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 flotend to draw the graph for analysis. We make use of logistic regression algorithm for prediction and it can be implemented using Apache Mahout Machine learning library. Then check the accuracy of the prediction.

4.2 Prediction Algorithm: Logistic Regression

Logistic Regression executes in two major phases:

- **Train the model**: This step is about creating a model using some train data, that can further be used for the classification of any input data rather I would say test data.

- **Test the model**: This step tests the generated model in step 1 by evaluating the results of classification of test data, and measuring the accuracy, scores and confusion matrix.

Supervised Learning algorithm is used for prediction because it provides most accuracy rather than traditional prediction technique.

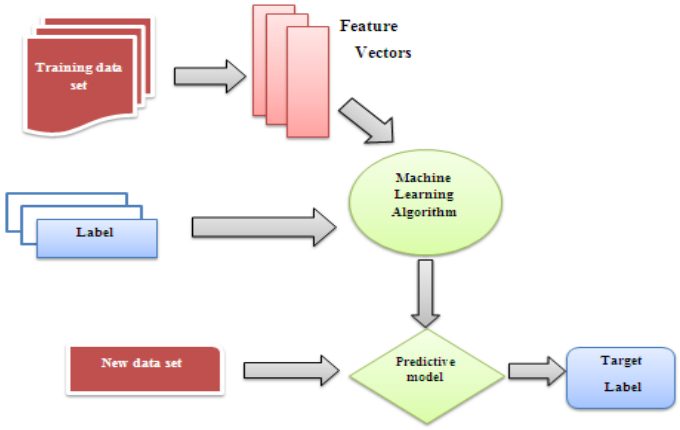

Figure: model of supervised learning algorithm

Logistic regression fits a special s-shaped curve by taking the linear regression, which could produce any y-value between minus infinity and plus infinity, and transforming it with the function:

$$p = Exp(y) / (1 + Exp(y))$$

which produces p-values between 0 (as y approaches minus infinity) and 1 (as y approaches plus infinity). This now becomes a special kind of non-linear regression, which is what this page performs.

Online-linear-regression (*D, number of iterations*)
Initialize weights w= (w_0, $w_{1, W2 \ldots. Wd}$)
for*i*=1:1*: number of iterations*
do **select** a data point from *Di=(xi, yi) from D*
seta=1/i
updateweight vector
w←w + a(y_i-f(x_i, W))x_i
end for
returnweights w

The advantages of using online linear regression algorithm is very easy to implement, continuous data streams,adapts to changes in the model over time

4.3 Establishment of Data Set

Experimental meteorological data includes number of properties namely site, period, monthly and annual mean of precipitation, temperature, cloud cover, vapor pressure, wet day frequency, diurnal temperature range and ground

ISSN: 2278 – 1323

frost frequency.In this work we concentrate on important three factors temperature, precipitation and cloud cover which are influences on most commonly visualized crops in Karnataka state. The parameters considered are shown in table 1.

Table 1. Hbase Database Table

Row key	Time-stamp	Column Family			
		Precipitation	Temperature		Cloud cover
Attribute	T	AP	Min T	Max T	AC

In this table 1, AP as average precipitation, MinT as minimum temperature, MaxT as maximum temperature, AC as average cloud cover. Attribute as a database table Rowkey, Timestamp is automatically assigned when there are in Hbase [10] .Precipitation, Temperature, cloud cover are clusters of three columns. Under each column there are sub columns is illustrated in table shown above.

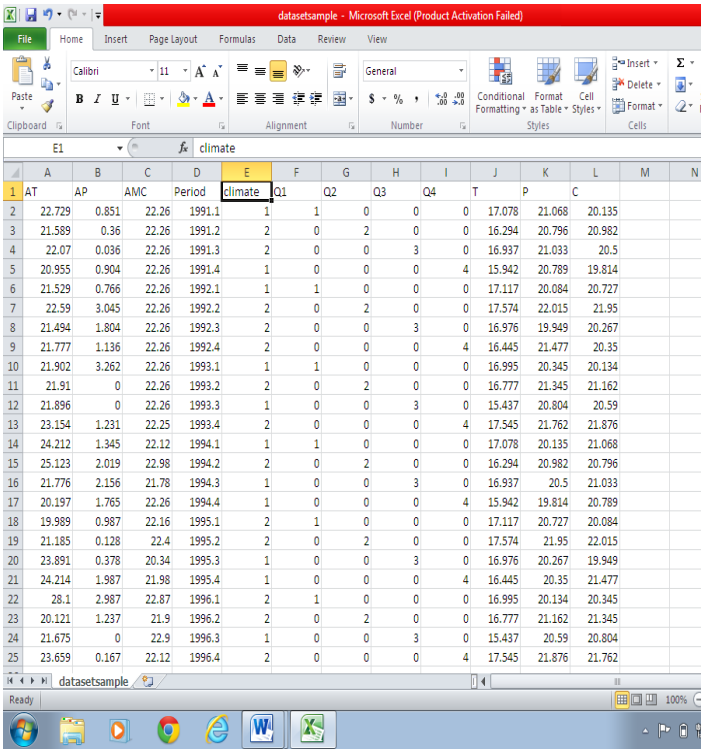

Figure 4: Sample weather data set for prediction

Here we construct a sample cleaned data set for giving as input to algorithm as shown in figure 4. In this datasets annual mean Temperature(AT), Annual mean Precipitation(AP), Annual mean Cloud cover(AMC), Period as input parameters, Quarter(Q1-Q4), Temperature (T), Precipitation (P), cloud cover (C) as predictive parameters. We make use climate as target variable. This table is made up with past 10 year's data items. Here we represent a prediction technique using Online Logistic Regression algorithm is running on apache mahout on top of hadoop. These supervised learning methods provide higher accuracy than statistical techniques.

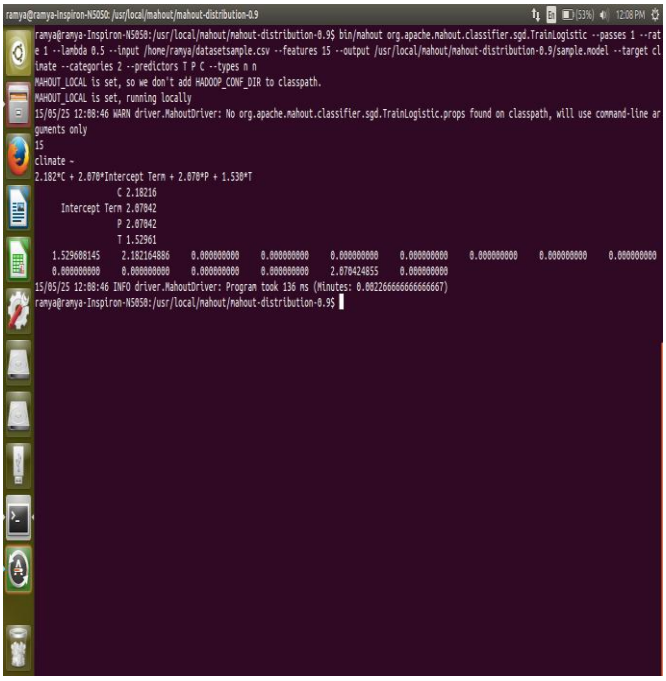

Figure 5: Execute Logistic regression algorithm to create Training model

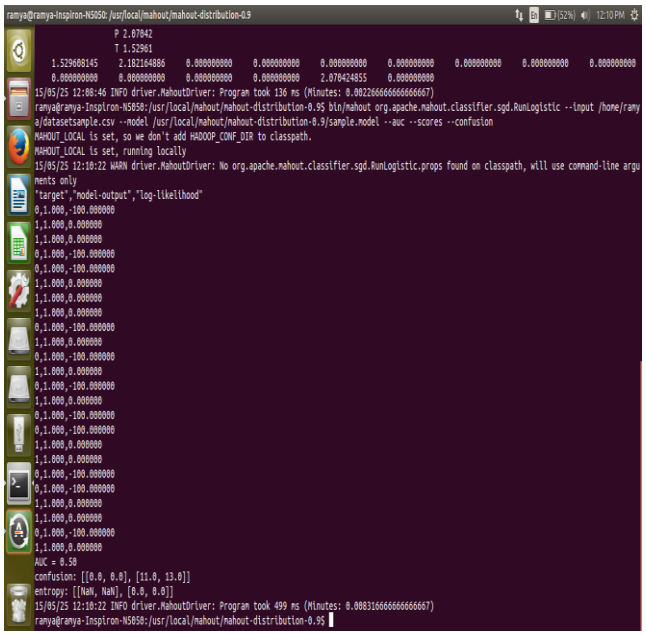

Figure 6: Accuracy analysis using testing model

Figure 5 and 6 shows the execution process of logistic regression algorithm on mahout. First command we run to create a training model of the sample data then we apply a testing model to check the accuracy of the algorithm, result is shown in the figures above.

V. RESULT ANALYSIS

This article adopts analysis of precipitation and temperature data set of Tumkur district region over the past 100 years (1901-2000) to conduct the forecasting

ISSN: 2278 – 1323

experiment and predict the data of the coming next month or year in advance, and uses the above weather prediction system which enables the logistic regression algorithm 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 flotend code to generate a graph. Here we analyze past 100 years weather data of Tumkur region and result will be produced. In figure 7 we analyze temperature data sets and calculate year wise and monthly wise average calculation for Tumkur region.

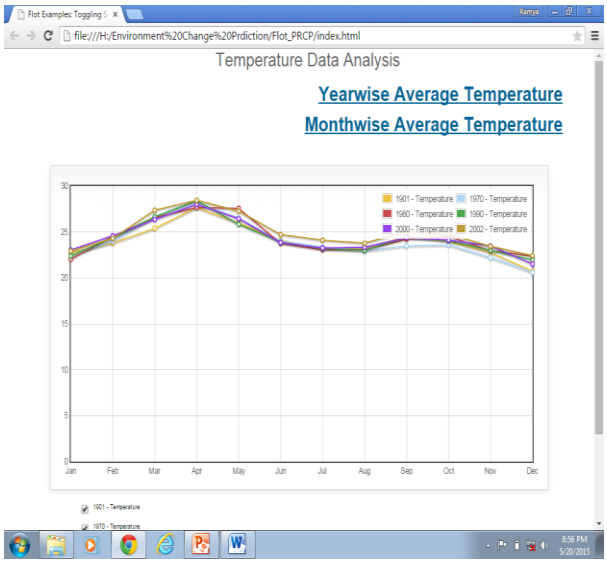

Figure 7: Analysis of annual and monthlytemperature data of Tumkur region

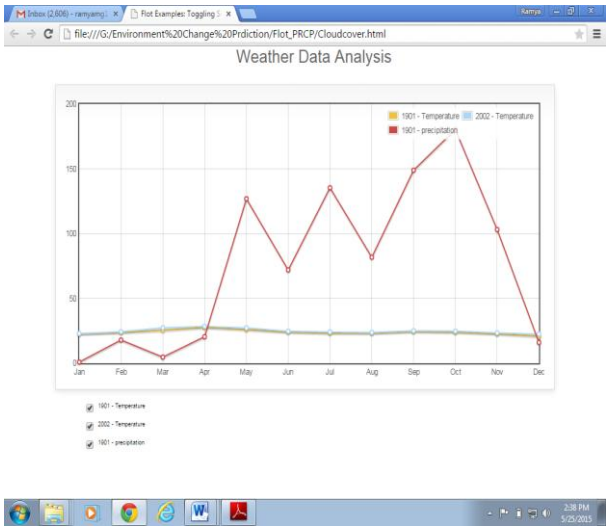

Figure 8: weather data analysis for Tumkur region

Figure 8 shows the analysis of different climate variables like maximum temperature, average temperature, minimum temperature, precipitation, cloud cover. Here analysis is based on historical data only for Tumkur region.

Table 2 defines the different parameters used as an input variable to run an algorithm. Pass is the number of times to pass over input data. Lambda is the amount of coefficient decay to use. Rate indicates Learning rate. Accuracy will be depending on these parameters. Table below shows accuracy increases while changing the corresponding variable value.

No. of Iterations	Input Parameters				
	Pass	Rate	Lambda	Runtime (ms)	Accuracy
01	1	1	0.5	183	0.5
02	50	10	0.5	870	0.72
03	100	50	0.5	729	0.76
04	100	50	0.7	744	0.78
05	200	60	0.7	990	0.84

Table 2. Accuracy Testing

VI. CONCLUSION AND FUTURE WORK

This paper designs the meteorological data storage as well as analysis platform based on Hadoop framework with the help of online logistic regression algorithm for prediction. This platform is based on distributed filesystem HDFS which incorporates distributed database Hbase, data warehouse management and efficient query processing tool Hive, data migration tool sqoop. The best data mining prediction algorithm regression also integrated into the system. This architecture has an ability of mass storage of meteorological data, efficient query and analysis, climate change prediction. We proposed a prediction technic with high accuracy. In future enhancement, later this prediction application can be used in different applications like emergency communication alerts in natural hazards like Cyclone, earthquake, tsunami.

ACKNOWLEDGMENTS

We are grateful to express sincere thanks to our faculties who gave support and special thanks to our department for providing facilities that were offered to us for carrying out this project.

REFERENCES

[1] Hai li and Xin Lu, "Challenges and trends of Big Data Analytics",Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing,2014
[2] Paras and Sanjay Mathur, " A Simple Weather Forecasting Model Using Mathematical Regression", Indian Research Journal of Extension Education Special Issue (Volume I), January, 2012
[3] Y.Radhika and M.Shashi, "Atmospheric Temperature Prediction using Support Vector Machines", international Journal of Computer Theory and Engineering, Vol. 1, No. 1, April 2009
[4] R. Usha Rani1, Dr. T.K Rama Krishna Rao, "An Enhanced Support Vector Regression Model for

ISSN: 2278 – 1323

Weather Forecasting", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 12, Issue 2, PP 21-24, May. - Jun. 2013

[5] Meonghun Lee, Jeonghwan Hwang and Hyun Yoe, "Agricultural Production System based on IoT", IEEE 16th International Conference on Computational Science and Engineering, 2013

[6] Reza Septiawan, AmrullahKomaruddin, Budi,"Prediction Model for Chilli Productivity Based on Climate and Productivity Data", UKSim-AMSS 6th European Modelling Symposium, 2012

[7] David Corne, Manjula Dissanayake and Andrew Peacock, "Accurate Localized Short Term Weather Prediction for Renewables Planning", IEEE, 978-1-4799-4546-7/14, 2014

[8] Seokhwan Yang and Jaechun Kim "A Prediction Model based on Big Data Analysis Using Hybrid FCM Clustering" The 9th International Conference for Internet Technology and Secured Transactions, 2014

[9] Duncan Waga, KefaRabah, "Environmental Conditions' Big Data Management and Cloud Computing Analytics for Sustainable Agriculture", World Journal of Computer Application and Technology 2(3): 73-81, 2014

[10] Leixiao Li, Zhiqiang Ma and Limin Li, "Hadoop-based ARIMA Algorithm and its Application in Weather Forecast", International Journal of Database Theory and Application Vol.6, No.5 (2013), pp.119-132 http://dx.doi.org/10.14257/ijdta.2013.6.5.11 ISSN: 2005-4270, 2013

ISSN: 2278 – 1323