VISVESVARAYA TECHNOLOGICAL UNIVERSITY,

JNANASANGAMA, BELAGAVI – 590018

Synopsis Report on

"CROP YIELD PREDICTION IN AGRICULTURE SECTOR USING MACHINE LEARNING"

Submitted in partial fulfillment of the requirement for the award of degree of

BACHELOR OF ENGINEERING

In

Computer Science and Engineering

Submitted by

QURATHFARHEEN A	[4MO18CS028]
DIVYA N	[4MO18CS010]
SUSHMITHA C	[4MO18CS035]
GEETHANJALI M V	[4MO18CS011]

Under the Guidance of

Shwetha M R

Asst. professor

Department of Computer science

Mysore College Of Engineering & Management, Mysuru.

Department of computer science Engineering

Mysore College Of Engineering & Management, Mysuru

#1072, T Narsipura Road, Near Big Banyan Tree Chikkhalli,

Mysuru, Karnataka 570028 2021-2022

ABSTRACT

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality ,pH ,EC,N,P,K etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data science. Through data science we extract the knowledge from the huge size of data. This system presents the study about the various data science techniques used for predicting the crop yield. The success of any crop yield prediction system heavily relies on how accurately the features have been extracted and how appropriately classifiers have been employed. This system summarizes the results obtained by various algorithms which are being used by various authors for crop yield prediction, with their accuracy and recommendation.

INTRODUCTION

India is an agriculture based country where most of the people derive their living from this sector. Agriculture is having a great impact on the country's economy. In the last decade India has seen serious natural calamities like drought or flood. Due to such disasters there is a huge loss to crop production and ultimately to the farmers. Due to such financial loss many farmers are committing suicide. If natural calamities are not present then there may be sudden pest attack destroying the crop. In any case farmer and the crop are always at the edge of risk. Government policies are there but that is not sufficient. Figure 1. shows the major crop producing states of India. Prediction of crop yield in advance can help the farmers and the Government bodies to plan for storage, selling, fixing minimum support price, importing /exporting etc.

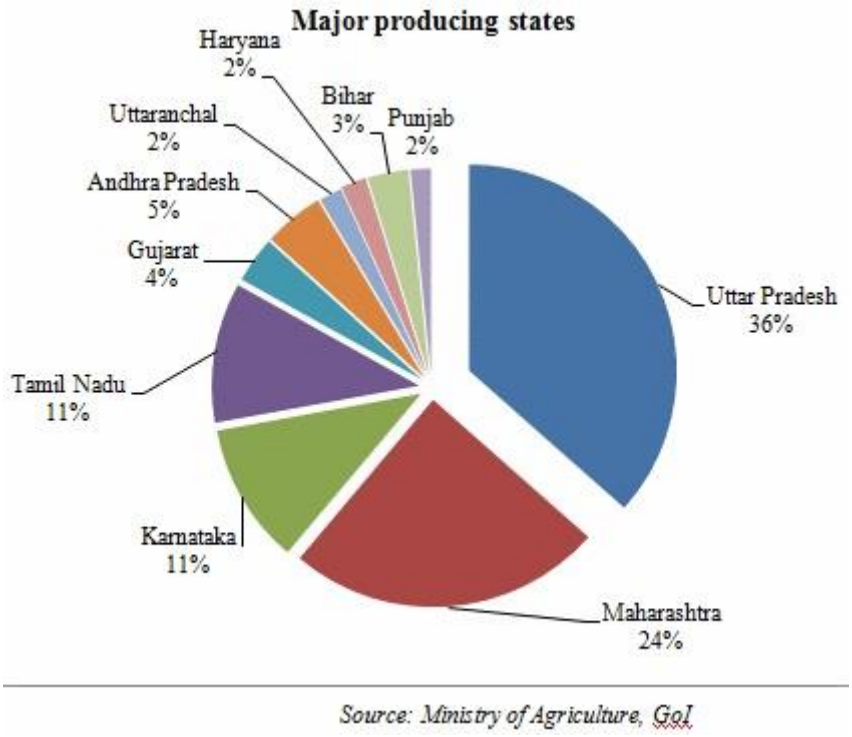


Fig.1: Major Producing States

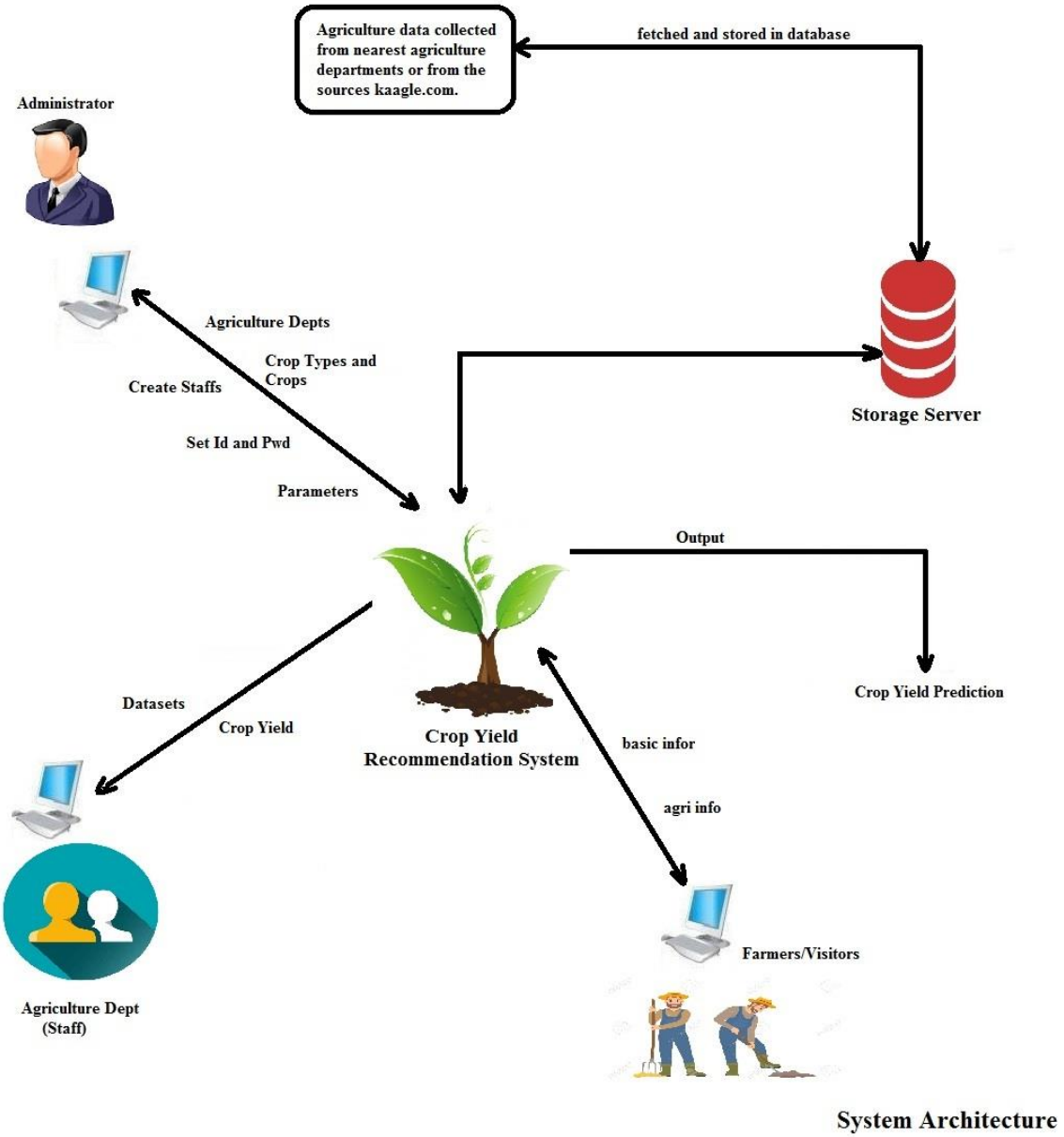


Fig: System Architecture

Till now only the past experience of the farmer was used which consist of randomly counting the number of seed buds that a plant is having and within each seed bud the number of seed it holds. Then based on experience the farmer used to predict the crop yield. Information technology can be used to avert the risk associated with the agriculture and it can also be used to predict the crop yield more accurately prior to harvest Yield prediction needs different kinds of data gathered from different sources like meteorological data, agri-meteorological, soil (pH, N, P, K) data, remotely sensed data, agricultural statistics etc. To handle such a huge data the best option we have is Data science. Data science is a method by which one can extract the knowledge from the huge bulk of data.

PROBLEM STATEMENT

Crop yield prediction based on the effect of temperature, rainfall, humidity and other constraints is an essential factor to deal with storage of crops. Current system is a manual prediction which requires more experience and time. Predicting the crop yield with high accuracy is one of the challenging task in the current agriculture sector.

EXISTING SYSTEM

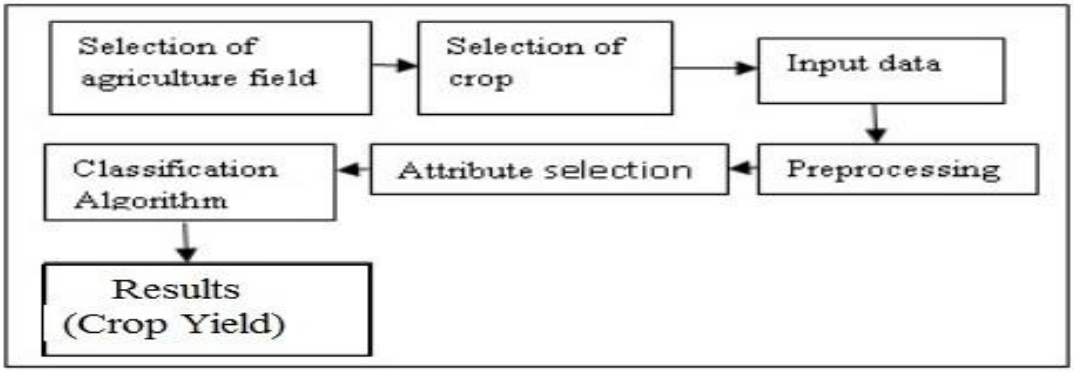
Prediction of agriculture yield is essential to deal with storage of crops, transportation decisions and risk management issues related to crops. Current system is manual where we compare the previous results with the present. Based on the previous experiences and results we come to know how much crop yield will be produced. There is no automation to predict the crop yield using the constraints temperature, rainfall, humidity, area and region.

Limitations

- Manual Process
- Time Consuming
- Less Reliable
- Less Efficient
- Less User Satisfaction

PROPOSED SYSTEM

Proposed system is an agriculture application which analyzes the previous data related to rainfall, temperature, humidity and crop yield. Proposed system makes use of data science in agriculture for decision making. The research is conducted taking under consideration the various constraints such as temperature, rainfall, humidity, region, area and other constraints. System uses "*Classification Rules*" technique - *Naive Bayes Algorithm* for crop yield prediction. Data collected from government sector (agriculture department) can be used to predict crop yield.



Brief overview of the Crop Yield Prediction System

1. Selection of agriculture field: Consider any agriculture field for the crop yield prediction system.

2. Selection of crop: consider any crop of choice which will be sown in that field.

3. Input data: Data may include information regarding soil (Nitrogen (N), Phosphorus(P), Potassium(K) content), Micro-nutrients present in soil, Moisture in soil etc which is collected over some period of time.

4. Pre-processing: Data which is collected should be preprocessed redundant data, inconsistent should be taken care.

5. Attribute Selection: Important Features have to be extracted.

6. Classification Algorithm: An appropriate and efficient algorithm should be employed.

7. Result: prediction or recommendation can be provided to the farmers based on the results obtained.

Scope and Objectives

- System objective is to estimate or forecast the yield of crop.
- System meant for multiple regions.
- System makes use of data collected from agriculture department.
- Proposed system makes use of data science in agriculture for decision making.
- System uses "*Classification Rules*" technique - Naive Bayes Algorithm for crop yield prediction.
- System makes use of SQL Serve to store the previous agriculture data.
- System works for dynamic data and System is a agriculture application useful for farmers.
- System improvises the farmers profits.

LITERATURE SURVEY

	Title	Author	Year of publication	Method Used	Result	Remarks
1.	Developing innovative applications in agriculture using data mining.	Sally Jo Cunningham and Geoffrey Holmes	1999	Weka classifier: 1.ZeroR 2.OneR 3.Naive Bayes 4.Decision-Table 5.Ibk 6.J48 7.SMO 8.Linear-Regression 9.M5Prime 10.LWR 11.Decision Stump. Association Rules: 1.Apriori Algorithm Clustering: 1. EM Clustering Algorithm.	The output from this is, literally, a classifier usually in the form of a decision tree or set of rules that can be used to predict the classificati on of a new data instance.	As the technology of machine learning continues to develop and mature, learning algorithms need to be brought to the desktops of people who work with data and understand the application domain from which it arises. Weka is a significant step in the transfer of machine

						learning technology into the workplace.
2.	Data Mining Techniques and Applications to Agricultural Yield Data.	D Ramesh, B Vishnu Vardhan.	2013	k nearest neighbor, k means algorithm.	In this process, given the rainfall in a specific year the system is in a position to predict the average yield production by considering the cluster in which the estimated rainfall belongs to.	The K-Means algorithm is able to partition the samples in clusters, but no considerations are made on the compounds that are responsible for this partition. Bi-clustering can provide this kind of information.
3.	Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach.	Monali Paul, Santosh K. Vishwakarma, Ashok Verma.	2015	K-Nearest Neighbor (KNN) and Naive Bayes (NB).	By the results we can see that the category having maximum confidence value is predicted as the category of that particular soil.	This study can help the soil analysts and farmers to decide sowing in which land may result in better crop production.

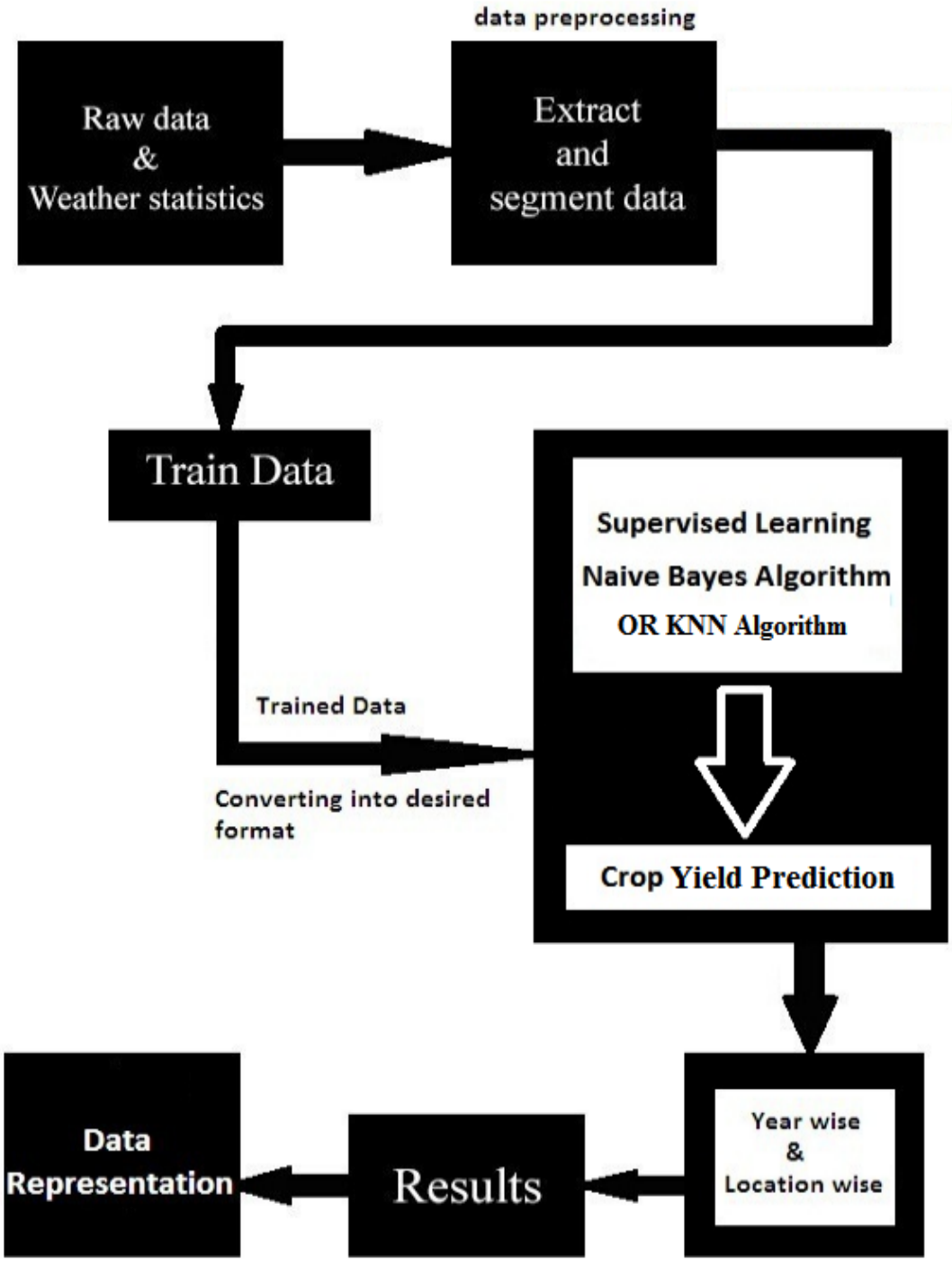
4.	Brief Survey of data mining Techniques Applied to applications of Agriculture.	Ami Mistry, Vinita Shah	2015	Classification technique: 1.Linear Regression 2. K - nearest neighbour. 3. Regression Tree 4. Support Vector Machine. Clustering Technique: 1.K- means clustering. 2. Self organised maps. 3. Density based clustering. Weight based clustering.	The results indicate that sunshine hours and daily temperatur e range play critical roles in rice yield variability in the current study area.	Farmer could plant different crops in different districts based on simple predictions made by this research and if that does take into effect, each and every farmer would get a chance at increasing their profits and increasing the country's overall produce. This will enable to have a better predictive model with more

						accurate results.
5.	A Study on Various Data Mining Techniques for Crop Yield Prediction.	Yogesh Gandge, Sandhya	2017	Classification Algorithm.	The output is the crop yield prediction per acre with some recommendation.	It is observed that the algorithm which is used by most of the authors does not uses a unified approach where in all the factors affecting the crop yield can be utilized simultaneously for predicting the crop yield.
6.	Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.	Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty.	2017	k-Nearest Neighbor, Decision Tree algorithm, ID3(Iterative Dichotomis) algorithm.	The result shows that Decision Tree Learning-ID3 algorithm gives a less value for percentage error than the KNN algorithm without omitting the outliers of the dataset.	Though the research is limited to some fixed dataset, the future ahead promises addition of more data that can be analysed with more

						machine learning techniques to generate crop predictions with better precision.
7.	Rice Yield Prediction Model Using Data Mining.	Umid Kumar Dey, Abdullah Hasan Masud, Mohamme d Nazim Uddin	2017	k means algorithm, Multiple Linear Regression, Algorithm, SVM Regression, Modified Nonlinear regression.	It is observed that it does an admirable job by predicting the yield with SVM regression providing best values.	It could be concluded that using modified Nonlinear Regression equations works better than the other three predefined models. It also proves that the MNR equation is the best fit.
8.	Effect of Temperatu re and Rainfall on Paddy Yield	Kuljit Kaur, Kanwalpre et Singh Attwal	2017	Apriori Algorithm.	The result shows that it predicts the growth of paddy yield. It	With increase in Rainfall the paddy yield also increased. During Reproducti ve phase

	using Data Mining				depends on various parameters such as Rainfall and Temperatu re.	the rainfall and temperatur e did not influence. During maturation phase, paddy yield was better expected at lower temp. and worse at high temp. Paddy yield was found to be high at low rainfall and low during high rainfall.

METHODOLOGY



Description

Step 1 : Raw data and Weather Statistics

This is the first step in the crop recommendation process where we collect agriculture data. Agriculture data collected from the region “mysore” which contains agriculture parameters, crop details, farmers details and parameter details. Agriculture parameters includes rainfall, temperature, soil features such as PH, nitrogen, potassium, iron etc…

Step 2: Extract and Segment Data (Data Preprocessing)

Here agriculture data analyzed and only relevant data extracted. The data required for processing extracted and segmented according to the different regions. Required data extraction is done because entire agriculture data not required for processing and if we input all data, it requires too much of time for processing, so data processing is done.

Step 3: Train Data

Once required data extracted and segmented, we need to train the data, train means converting the data into the required format such as numerical values or binary or string etc.. conversion depends on the algorithm type.

Step 4: Supervised Learning

ML concerns with construction and study of system that can learn from data.

For example, ML can be used in E-mail message to learn how to distinguish between spam and inbox messages.

Supervised learning is an approach to machine learning that is based on training data that includes expected answers.

Naive Bayes Algorithm OR KNN Algorithm

"*Naive Bayes Algorithm*" is used for crop recommendation because of the following reasons;

1. efficient classifier

2. works fine for less number of parameters as well as more number of parameters.

3. Works fine for small data-set as well as big data-set.

4. more accurate results

Step 5: Crop Yield Prediction

System predicts the yield for the selected crop based on the agriculture parameters using machine learning algorithm.

Step 6: Location and Year Based

The crop yield prediction is done based on the region wise as well as year wise.

Step 7: Results

Here we find the accuracy of the algorithm by dividing the training datasets into training and testing datasets. 90% considered as training datasets and 10% considered as testing datasets.

Step 8: Visual Representation

Crops yield displayed for the farmers on GUI. When users gets login to the application system recommends suitable and high profit crops for the farmers on a GUI.

SYSTEM REQUIREMENTS

1. Software Requirements:

- IDE : **Visual Studio**
- Programming Language : **C#**
- Back End : **MS SQL Server**

2. Hardware Requirements

- RAM : **2GB+**
- Processor : **Pentium 4+**
- Processor Speed : **2ghz+**

EXPECTED OUTPUT

Outcome of the Project

System uses different agriculture parameters for yield prediction, parameters such as temperature, rainfall, PH value, nitrogen, potassium, zinc, iron etc. Training datasets trained using ML algorithm "bayesian classifier" and crop yield for future will be predicted in tons.

Input and Output

- **Input –** Previous year's agriculture data which includes temperature, rainfall, humidity and other constraints.
- **Output –** predicts crop yield using different constraints such as region, temperature, rain, humidity and yield prediction based on year wise and location wise.

REFERENCES

[1] Maria Rossana C.de Leon,Eugene Rex L.Jalao, "A prediction model framework for crop yield prediction," Asia Pacific Industrial Engineering and Management System.

[2] Ramesh A. Medar,Vijay S. Rajpurohit, "A survey on Data Mining Techniques for crop yield prediction", International Journal of Advance Research in Computer Science and Management Studies.Volume 2 , Issue 9, Sept 2014.

[3] D Ramesh, B Vishnu Vardhan, "Region specific crop yield Analysis :A data Mining approach ",UACEE International Journal of Advancees in computer Science and its Applications-IJCSIA volume 3:issue 2.

[4] S.Veenadhari, Dr Bharat Mishra, Dr C D Singh, "Soyabean productivity modeling using Decision Tree Algorithms," International Journal of Computer Applications Vol 27-No7 Aug 2011.

[5] George RuB, "Data Mining of Agricultural Yield Data:A comparison of Regression models.

[6] Wim G.M. Bastiaanssen, Samia Ali, "A new crop yield forecasting model based on satellite measurements applied across the Indus Basin,Pakistan," Agriculture, Ecosystems and Environment 94 (2003) 321–340.

[7] Sudhanshu Sekhar Panda, Daniel P. Ames , and Suranjan Panigrahi, "Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques" , Remote Sensing 2010, 2, 673-696; doi:10.3390/rs2030673

[8] S.Veenadhari,Dr. Bharat Misra ,Dr. CD Singh," Machine Learning Approach for forecasting crop yield based on climatic parameters", 978-1- 4799-2352-6/14/$31.00 ©2014 IEEE

[9] Jeferson Lobato Fernandes; Jansle Vieira Rocha," Sugarcane yield estimates using time series analysis of spot vegetation images", Rubens Augusto Camargo Lamparelli, Sci. Agric. (Piracicaba, Braz.), v.68, n.2, p.139-146, March/April 2011

[10] Yunous Vagh, Jitian Xiao," Mining temperature profile data for shire-level crop yield prediction", 978-1-4673-1487-9/12/$31.00 ©2012 IEEE, Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

[11] M. Gunasundari, Dr. T Arunkumar, Ms. R. Hemavathy," CRY – An improved Crop Yield Prediction model using Bee Hive Clustering 423 Approach for Agricultural data sets", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22,978-1-4673-5845-3/13/$31.00©2013 IEEE

[12] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman," Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh", 978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE SNPD 2015, June 1-3 2015, Takamatsu, Japan

[13] R. Sujatha, Dr. P.Isakki,A Study on Crop Yield Forecasting Using Classification Techniques, 978-1-4673- 8437-7/16/$31.00 ©2016 IEEE.

[14] Niketa Gandhi, Leisa Armstrong," Applying Data Mining Techniques to predict yield of Rice in Humid Subtropical Climatic Zone of India". 978-9-3805-44212/16/$31.00_c 2016 IEEE.

[15] Monali Paul, Santosh K. Vishwakarma, Ashok Verma," Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" ,2015 International Conference on Computational Intelligence and Communication Networks, 978-1-5090-0076-0/15 $31.00 © 2015 IEEE, DOI 10.1109/CICN.2015.156