A Survey on Crop Yield Prediction using Data Mining

[1]Mythra.N, [2]Dr.Velayudham.A, [3]Dr.Shamila.E.S, [4]Pavithra.M

[1]PG Scholar, [2]Professor& Head, [3]Associate Professor,[4]Assistant professorDepartment of Computer Science and Engineering, Jansons Institute of Technology, Coimbatore, Tamil Nadu, India

Abstract

Now-a –days the most important emerging field in the real world is agriculture and it is the main occupation and backbone of our country. The most of the primary and secondary business is based on farming and it plays the main role in our Indian economy. In farming, season cultivation of the crop is based on the soil. Another important work in agriculture is selecting soil is based on the weather condition and also when to harvest the crop for the best cultivation. Data mining techniques is very popular in the Area of agriculture. Data mining is the process of finding the new templates from the large data sets there are various data mining techniques are used and evaluated in agriculture for estimating the future year's crop production. Utilization of information and communications technology enables automation of extracting significant data in an effort to obtain knowledge and trends crop is an essential key factor of agriculture. So data mining techniques are useful for new framers. This paper presents a brief comparative study of various papers that deals with various techniques used to figure out the crop yield different data mining techniques that are in the use for crop yield estimation in agriculture. Recently data processing techniques in agriculture required decide, storage, monitored and retrieval the resource used this survey aims to search out appropriate data processing models to realize high accuracy and prediction capabilities.

Keywords - *Data Mining, Agriculture, framer, Crop Selection, soil selection, methodology, high accuracy, prediction capabilities.*

I. INTRODUCTION

Data Mining is the process of extracting useful information and patterns from enormous data. Data Mining includes collection, extraction, analysis and statistics of data. It is also known as Knowledge discovery process, Knowledge Mining from Data or data/ pattern analysis. Data Mining is a logical process of finding useful information to find out useful data. Once the information and patterns are found it can be used to make decisions for developing the business. Data mining tools can give answers to your various questions related to your business which was too difficult to resolve.

They also forecast the future trends which lets the business people to make proactive decisions. Data mining involves three steps. They are

- ***Exploration*** – In exploration the data is cleared and converted into another form. The nature of data is also determined.
- ***Pattern Identification*** – In this step is to choose the pattern which will make the best prediction.
- ***Deployment*** – The identified patterns are used to get the desired outcome.

A. Benefits of Data Mining
- Automated prediction of trends and behaviours.
- It can be implemented on new systems as well as existing platforms.
- It can analyse huge database in minutes.
- Automated discovery of hidden patterns.
- There are a lot of models available to understand complex data easily.
- It is of high speed which makes it easy for the users to analyse huge amount of data in less time.
- It yields improved predictions.

B. Data Mining Techniques
One of the most important task in Data Mining is to select the correct data mining technique. Data mining technique has to be chosen based on the type of business and the type of problem your business faces. A generalized approach has to be used to improve the accuracy and cost effectiveness of using data mining techniques. There are basically seven main Data Mining techniques which is discussed in this article. There are also a lot of other Data Mining techniques but these seven are considered more frequently used by business people.

- Statistics.
- Clustering.
- Visualization.
- Decision Tree.
- Association Rules.
- Neural Networks.
- Classification.

1. *Statistical Techniques*

Data mining techniques statistics is a branch of mathematics which relates to the collection and description of data. Statistical technique is not considered as a data mining technique by many analysts. But still it helps to discover the patterns and build predictive models. For this reason data analyst should possess some knowledge about the different statistical techniques. In today's world people have to deal with large amount of data and derive important patterns from it.

2. *Clustering*

Clustering is one among the oldest techniques used in Data Mining. Clustering analysis is the process of identifying data that are similar to each other. This will help to understand the differences and similarities between the data. This is sometimes called segmentation and helps the users to understand what is going on within the database. For example, an insurance company can group its customers based on their income, age, nature of policy and type of claims.

There are different types of clustering methods. They are as follows

- Partitioning Methods
- Hierarchical Agglomerative methods
- Density Based Methods
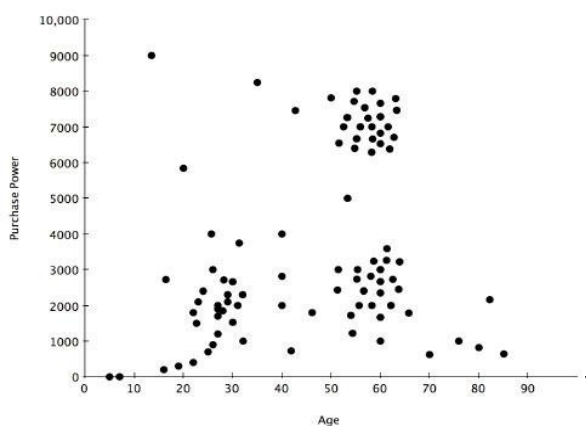- Grid Based Methods
- Model Based Methods

Figure 1: Clustering.

3. *Visualization*

Visualization is the most useful technique which is used to discover data patterns. This technique is used at the beginning of the Data Mining process. Many researches are going on these days to produce interesting projection of databases, which is called Projection Pursuit. There are a lot of data mining technique which will produce useful patterns for good data. But visualization is a technique which converts Poor data into good data letting

different kinds of Data Mining methods to be used in discovering hidden patterns.

4. *Induction Decision Tree Technique*

A decision tree is a predictive model and the name itself implies that it looks like a tree. In this technique, each branch of the tree is viewed as a classification question and the leaves of the trees are considered as partitions of the dataset related to that particular classification. This technique can be used for exploration analysis, data pre-processing and prediction work. Decision tree can be considered as a segmentation of the original dataset where segmentation is done for a particular reason. Each data that comes under a segment has some similarities in their information being predicted. Decision trees provides results that can be easily understood by the user. Decision tree technique is mostly used by statisticians to find out which database is more related to the problem of the business. Decision tree technique can be used for Prediction and Data pre-processing.

6. *Neural Network*

Neural Network is another important technique used by people these days. This technique is most often used in the starting stages of the data mining technology. Artificial neural network was formed out of the community of Artificial intelligence. A neural network is a collection of interconnected neurons. Which could form a single layer or multiple layer. The formation of neurons and their interconnections are called architecture of the network.

7. *Association Rule Technique*

This technique helps to find the association between two or more items. It helps to know the relations between the different variables in databases. It discovers the hidden patterns in the data sets which is used to identify the variables and the frequent occurrence of different variables that appear with the highest frequencies.

8. *Classification*

Data mining techniques classification is the most commonly used data mining technique which contains a set of pre classified samples to create a model which can classify the large set of data. This technique helps in deriving important information about data and metadata (data about data). This technique is closely related to cluster analysis technique and it uses decision tree or neural network system. There are two main processes involved in this technique,

- ➢ Learning – In this process the data are analysed by classification algorithm

➢ Classification – In this process the data is used to measure the precision of the classification rules

C. Agriculture

Agriculture is the cultivation of land and breeding of animals and plants to provide food, fibre, medicinal plants and other products to sustain and enhance life. Agriculture was the key development in the rise of sedentary human civilization, whereby farming of domesticated species created food surpluses that enabled people to live in cities. The study of agriculture is known as agricultural science. The history of agriculture dates back thousands of years; people gathered wild grains at least 105,000 years ago and began to plant them around 11,500 years ago before they became domesticated. Pigs, sheep, and cattle were domesticated over 10,000 years ago. Crops originate from at least 11 regions of the world. Industrial agriculture based on large-scale monoculture has in the past century come to dominate agricultural output, though about 2 billion people worldwide still depend on subsistence agriculture.

D. Crop Prediction

Crop prediction is major task done in agriculture field for high productivity. Crop alteration has been practiced by humankind for thousands of years, since the beginning of civilization. Altering crops through breeding practices changes the genetic make-up of a plant to develop crops with more beneficial characteristics for humanscrop is main source for the framing.

There are nine types of framing
1. Subsistence framing
2. Shifting agriculture
3. Plantation agriculture
4. Intensive framing
5. Dry agriculture
6. Mixed and multiple agriculture
7. Crop rotation
8. Sedentary cultivation or permanent agriculture
9. Terrace cultivation

1. Subsistence framing

Subsistence farming is useful for the poor framers .majority of framers in the country practise subsistence framing the whole family works on the farm the follow the tradition methods dry land framing is practised here the crops like jowar,bajra,and pluses need less water the result of the yield is not very high.

2.Shifting agriculture

It is practised in the forest area by cutting and burning the trees seeds are planted in the ground it does not involve in ploughing the soil this cultivation leads to soil erosion and loss of soil fertility .This is a wasteful method of cultivation.

3.Plantation agriculture

Tree framing is known as plantation framing introduced by British in nineteenth century some of the plantation like tea, coffee and rubber and it is practised in the hill areas.

4.Intensive framing

This type of framing is based on use of large-scale fertiliser and pesticides. Mechanised agriculture by introducing machines in various process of framing this method practised in unitarea.

5. Dry agriculture

Crops are cultivate in area can withstand in dry condition generally the crop yield is low in these areas dry agriculture are practised in the dessert areas.

6. Mixed and multiple agriculture

Multiple framing refers to two or more crops together .In this case number of crops varying maturing periods at same time and this methods are practised in hill area were good rainfall is need for this type of cultivation.

7. Crop rotation

Crop rotation refer to crop one after another in fixed rotation to maintain the soil fertility this takes an year to complete selection of crop based on the soil.

8. Sedentary cultivation or permanent agriculture

It is also known as settled cultivation here framer get settled at the place and practice continued use of land year after year with the variation of crops permanent settlement of framer exists.

9. Terrace cultivation

This method is practised in area of hills and mountain slopes land are in slopping nature soil erosion is also checked due to terrace formation on hill slops.

II. LITERATURE SURVEY

A. Analysing Soil Data Using Data Mining Classification Techniques

V. Rajeswariand K. Arunesh [1] 2016 provides an review on data mining classification techniques to predict the soil type soil is an essential key factors of agriculture The objective of the work is to predict soil type using data mining classification techniques .soil type plays Important role in agriculture The variety of association techniques in data mining are applied in data base that useful to predict different soil type in framing

Methods and techniques used in agriculture data mining

1. JRip

This algorithm was proposed by William W. Cohen. IREP optimized version is Repeated Incremental Pruning to produce error Reduction (RIPPER) This algorithm is a propositional guideline learner .J-Rip classifier is one decision tree pruning models based on association rules This technique is best to reduce error pruning

2. J48 (C4.5)

This algorithm was proposed by Ross Quinlan the J48 is one of the classification -decision tree algorithm and it slightly modified from C4.5 in weka c4.5 is refers to statistical classifier .J48 predicts dependent variable from available data .It builds tree based on attributes values of training data this classification data with the help of feature of data instances that said to have information gain The method importance of error tolerance is developed using pruning concept

3. Naive Bayes

It is one classifiers in simple probabilistic classification techniques in machine learning it is based on Bayes theorem with independence features each class labels are estimated through the given instance and it need only small amount of training data to predict class necessary for classification

From the research analysis J48 classification algorithm provides high accuracy rate in soil fertility techniques used in this paper JRIP,J48, and Naïve Bayes These classification algorithms are applied to extract the knowledge from soil data and two types of soil considered such as red and black The JRIP model can produce accurate result among other models.

B. Green Farm –dm:a Tool for Analyzing Vegetable Crops Data from a Greenhouse using Data Mining Techniques (first trial)

K. L. Ponce-Guevara, J. A. Palacios-Echeverrıa, E. Maya-Olalla, H. M. Domınguez-Limaico,L. E. Suarez-Zambrano, P. D. Romero-Montalvo, D. H. Peluffo-Ord´o˜nez and J. C. Alvarado-Perez[2]2017 explained about Green Farm-DM is tool used for analysing the growth of the crops and moisture level in soil within a greenhouse using big data and data mining techniques its helps to predict the affects that cause the growth of the plants like relative humidity,soil moisture,ambient temperature ,level of illumination and co2 these parameters are essential foe photosynthesis results are classified in two stages first stage we use C4.5(Chosen for the first trial)produce graph with use of decision tree and second stage is prediction based on the pervious stage with help of predicted new dataset on soil.

1. Methods
a. Big data

Big data is term refers to either the information that exceeds the processing capacity of conventional database systems, information size that is so large and constantly changing or simply the fact it does not fit into the structure of database within such information, hidden patterns are valuable for decision making.

b. KDD process

This process involves some stages in the analysis of data selection, processing, transformation

c. Data mining

Data mining is the process of extracting implicit and potentially useful information from the data.

C. Analysis Of Crop Yeild Prediction Using Data Mining Techniques

D Ramesh, B Vishnu vardhan[3] 2015 presents an overview about the crop yield prediction using data mining techniques we can estimate the future year crop production in agriculture using two techniques multiple linear regression(MLR) and density based clustering techniques density based clustering needs many constants and less sensitive to outliers. Results of these methods are practised in district of Andhra Pradesh it help the framers to predict the specific crop.

1. Methods
a. Multiple linear regression

A regression model that involves more than one predictor variable is called multiple regression model linear relationship between dependent and independent variable. Dependent variable termed as predicant and independent variables are called predictors. This Multiple linear regression (MLR) techniques is based on the least squares and mostly used in climatology and also used for predicting the production.

b. Density –based clustering techniques

This method is used to group the objects according to specific density function Density is usually defined as the number of objects in a particular neighbourhood of data objects. In these approaches, a given cluster continues to grow as long as the number of objects in the neighbourhood which exceeds some parameter. This is considered to be different from the idea in partitioning algorithms that use iterative relocation of points that give a certain number of clusters.

D. Crop Yeild Prediction With Aid Of Optimal Neural Network In Spatial Data Mining: New Approaches

Aakunuri Manjula1 and Dr. G.Narsimha [4]2016 explained about Data mining techniques used in many sectors to predict the expected information Data mining techniques used in agriculture for crop yield to predict based on parameters like weather condition,diseases and pests ,planning of harvesting

and other biological factors etc. .to predict the healthy crop in future methods used in this paper is mutilinear principal component analysis(MPCA) and optimal neural network (ONN) problem defined time consuming , In existing time series crop yield prediction method does not react to variations that occur for cycles and seasonal effects. Needs extensive information to develop and test the model and also available information in agriculture is sparse and incomplete in existing simulation model. Limited studies have been made in crop yield prediction using existing decision tree technique. Prediction error value also important problem in crop yield prediction or estimation methods.

1. Problems definition

1) the most important problem of existing crop yield prediction method is accuracy and time consuming problem.

2) In existing time series crop yield prediction method does not react to variations that occur for cycles and seasonal effects.

3) Needs extensive information to develop and test the model and also available information in agriculture is sparse and incomplete in existing simulation model.

4) Limited studies have been made in crop yield prediction using existing decision tree technique.

5) Prediction error value also important problem in crop yield prediction or estimation methods.

6) These are the main drawbacks of various existing works, which motivate us to do this research on crop yield prediction.

E. Data mining and wireless sensor network for groundnut pest /disease interaction and predictions- a preliminary study

A. K. Tripathy, J. Adinarayana, D. Sudharsan, K. Vijayalakshmi , S. N. Merchant, U. B.Desai [5]2013 provides a study and analysis of pest (Thrips)/disease (Bud Necrosis) management for dynamic crop-weather data for precision agriculture forgroundnut (peanut) crop Data mining techniques are trained through mathematical models to predict the Bud Necrosis and thrips WSN was established in the test bed to obtain real-time. Weatherparameters the crop-weather-pest/disease dynamics and hidden. Relations were obtained and quantified using DM techniques. This will help the framer to take strategic decision to save the crop from pest/disease affects and improve yield in agriculture.
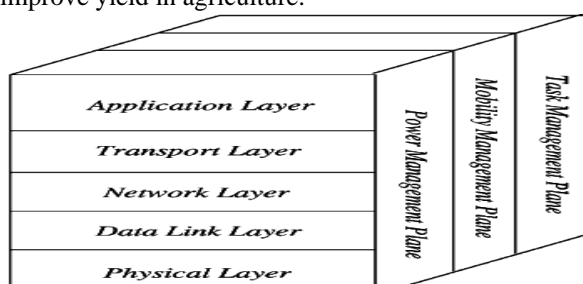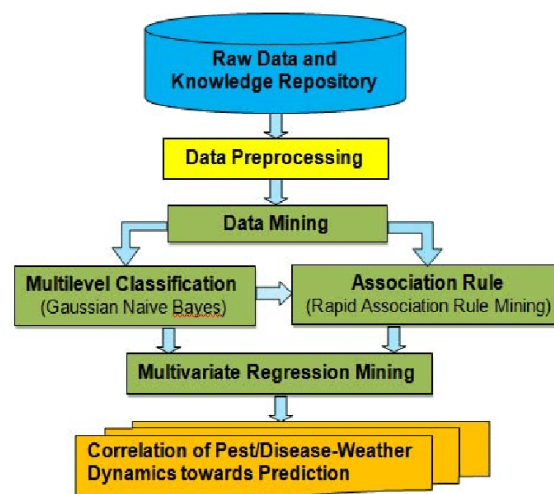

Figure3: WSN architecture

Figure 4:DM Processing flow for Pest/DiseaseDynamics.

F. Data Mining Techniques to Predict the Accuracy of the Soil Fertility

S. Hari Ganesh, Mrs. Jayasudha[6]2015 explained about Data mining plays important role in young research field. Data mining is very popular in area of agriculture. There many application software are available ,soil is the important resource for agriculture to test the soil fertility ,with use soil dataset information can predict the crop yield for the present season and soil testing and classification plays important role in fertility ,algorithms used navies bayes,j48(c4.5),JRIP this will helpful for the framer to recommend suitable fertilizer.

1. Methods
a. NaïveBayes

Naïve Bayes is one of the simple probabilistic classifier based on the applying Bayes theorem with strong independence assumptions.

2. ADVANTAGE

It only requires a small amount of training data to estimate the parameters (means and variance of the variables necessary for the classification.

3. JRip

This algorithm was proposed by William W. Cohen. IREP optimized version is Repeated Incremental Pruning to produce error Reduction (RIPPER) This algorithm is a propositional guideline learner .J-Rip classifier is one decision tree pruning models based on association rules This technique is best to reduce error pruning.

G. Application of data mining in agriculture

B. Milovi and V. Radojevi[7]2015 presents an agricultural organization is useful in various sector data mining techniques are mostly used in agriculture organization they collect the large amount of data for processing and retrieval data mining in addition about crop enables the agriculture enterprises predict the

customer behaviour based on differ perspective It provides the opportunity for exploring the hidden patterns in form of collection of data by data mining which helps to determine the condition of customer organization data mining techniques useful for decision making.

A. Methods
1) Classification and regression
2) Association rule
3) Cluster analysis

H. Different types of data mining techniques used in agriculture –a survey

R S. Kodeeshwari, K. Tamil Ilakkiya[8] 2017presents a survey paper on different data mining techniques in agriculture framing plays an important in role in Indian economy ,decision making is the important process for framer to improve the crop yield so data mining is the domain suitable for agriculture mining in agriculture is a innovate groundwork domain data mining techniques such as classification, association rule, regression these techniques help the framer for high production with supplementary profit.

Problems in data mining
- Progress a consolidated theory of data mining
- Maximize for large structural data and high speed data streams
- Mining time series data and ordered data
- Data mining has composite knowledge from complicated data
- Data mining has composite knowledge from complicated data
- Data mining in certain network environment
- Mining multi-agent data and scattered data mining
- Data mining for inorganic and atmospheric problems

I. A survey on various data mining techniques in field of agriculutre for prediction of crop yeild

Huma Khan, Shahista Navaz, Dr. S. M. Ghosh[9]2015provides a survey on prediction of crop yield in agriculture using data mining techniquesThe information of raw data in field of agriculture in state and geography wise of the crop yield it helps the framer for predict the best crop classification and clustering techniques are two data mining techniques used for better crop productivity and helps the framer for the decision making and provides the efficiency.

A. Methods
1) k-means, k-Nearest Neighbor (KNN)
2) support vector machine (SVM)
3) Multiple linear Regression (MLR)

J. Appliction of Data Mining Techniques for Prediction of Crop Production in India

Suvidha Jambekar, Shikha Nema ,Zia Saquib[10]2018 presents a survey on application of data mining in crop yield agriculture is main occupation of India .about 70% of primary and secondary business is based on framing India is top countries to produces the major crops like rice ,wheat and maize here data mining techniques is applied in agriculture sector to predict the crop yield weather condition play a major role for the healthy crop productivity .algorithms used in data mining techniques -multiple linear regression, random forest regression and support vector regression are used this help the framer to choose the crop and provides the efficient result.

Methods
1) Multiple linear regression
2) Random forest regression
3) Support vector regression

III. CONCLUSION

Agriculture plays the major role in developing countries like India. Using some new technology in framing can change the scenario of decision making and framers to yield crop production in better way. Data mining in the agriculture discipline can be perceived to be novel undertaking a practice that is targeted at improving the welfare of the people .every framer is highly interested with the level of yields to expected in given month or season earlier on prediction of crop yield was performed through analysis of the past experience by the framer .data mining procedures can be employed to perform prediction on possible future yields Data mining, through better management and data analysis in data mining with amount of available information it is easier to make the decision .analysis of our parameters give the impact to the agriculture , bye the use of data mining techniques acquired knowledge can be used to make successful decision which will advance the success of the agriculture organization on the market .data mining technology provides user oriented access to new and hidden patterns in data from which knowledge is generated which can help with decision making in agriculture organization use of data mining techniques in agriculture field creates condition for mankind adequate decision and with that achieving competing advantage .increased drought and floods are likely to increased production variability Data mining plays important role in decision making. In this paper we discussed about the role of data mining techniques in agriculture field and it is useful for the farmers for decision making about crop yield prediction. This paper combines the works of many authors and it useful for the current circumstances in the agriculture domain. The main aim of this survey, is to upgrade the procedure of data

mining techniques in framing .so that the framers get high production with supplementary profit.

REFERNCES

[1] Rajeswari and k.arunesh, "Analysing Soil Data using Data Mining classification techniques", Indian Journal of science and Technology,Volume 9(19), DOI: 10.17485/ijst/2016/v9i19/93873, May 2016.

[2] K. L. Ponce-Guevara1, J. A. Palacios-Echeverr´ıa1, E. Maya-Olalla1, H. M. Dom´ınguez-Limaico1,L. E. Su´arez-Zambrano1, P. D. Rosero-Montalvo1,4, D. H. Peluffo-Ord´o˜nez1,2, and J. C. Alvarado-P´erez2, "Green Farm-DM: A tool for analysing vegetable crops data from a greenhouse using data mining techniques(First trial)", 978-1-5386-3894-1/17/$31.00 ©2017IEE.

[3] D Ramesh, B Vishnu vardhan, "ANALYSIS OF CROP YEILD PREDICTION USING DATA MINING TECHNIQUES", IJRET: International Journal of Research in engineering and Technology, volume: 04, Issue: 01, jan-2015.

[4] Available @ http://www.ijret.org

[5] Aakunuri Manjula and Dr.G.Narsimha, "Crop Yield Predictionwith Aid of Optimal Neural Network in Spatial Data Mining: NewApproaches", InternationalJournal of Information & Computation Technology, ISSN 0974-2239 volume 6, Number1, pp. 25-33, 2016.

[6] © International Research Publications House http://www. irphouse.com.

[7] A. K. Tripathy, J. Adinarayana, D. Sudharsan, K. Vijayalakshmi , S. N. Merchant, U. B.Desai, "Data Mining and Wireless Sensor Network for Groundnut Pest/Disease Interaction and predictions –A Preliminary Study", International Journal of Computer Information Systems and Industrial Management Applications, ISSN 2150-7988 Volume 5,pp. 427-436,©2013.

[8] www.mirlabs.net/ijcisim/index.html.

[9] S. Hari Ganesh, Mrs. Jayasudha, "Data Mining Technique to Predict the Accuracy of the soil Fertility ",International Journal of Computer Science and Mobile Computing (IJCSMC),volume 4,Issue 7 ,pg.330-333,july 2015.

[10] B. Milovi and V. Radojevi, "APPLICATION OF DATA MINING IN AGRICULTURE", Bulgarian journal of agricultural science, 21(No 1) 26-34agricultural academy, 2015.

[11] R S. Kodeeshwari, K. Tamil, "Different Types of Data Mining Techniques Used in Agriculture –A Survey "International Journal of Advanced Engineering Research and Science (IJAERS), vol-4, Issue-6s, Jun-2017.

[12] https://dx.doi.org/10.22161/ijaers.4.6.3.

[13] Huma Khan, Shahista Navaz, Dr. S. M. Ghosh, "A Survey on Various Data Mining Techniques in Field of Agriculture for Prediction of Crop Yields", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96 | Impact Factor, 2015.

[14] Suvidha Jambekar, Shikha Nema ,Zia Saquib, "Data Mining Techniques for Prediction of Crop Production in India", International Journal of Innovations &Advancement in Computer Science(IJIACS),ISSN 2347 – 8616,Volume 7, Issue 4,April 2018.

[15] Ramesh A, Vijay. S. Rajpurohit, "A survey on Data Mining Techniques for Crop Yield Prediction", "International Journal of Advance Research in Volume 2, Issue 9, September 2014".

[16] Zekarais Dirida,Berhanu borena, "application of data mining techniques for crop productivity prediction", "HiLcoE journal of computer science and technology ''vol.1novmeber.2015

[17] Dakshayini Patil,Dr.M.S.shirdhonkar, "rice crop yield prediction using data mining techniques", "International journal of advanced research in computer science and software engineering "volume7,issue 5,may2017

[18] DR.A.Senthil Kumar,P.Arun, "A survey on agriculture for crop yield prediction using data mining techniques", "IOSR journal of computer engineering (ISOR-JCE)p.no23-25,vol 6,issue 3,March 2017

[19] Yogesh gandge,sandhya, "A study on various data mining techniques for crop yield prediction, , " international conference on computer and optimization techniques (ICEECCOT), 978-1-5386-2361-9/17/$31.00 ©2017 IEEE