Design And Implementation Of Crop Yield Prediction Model In Agriculture

Sangeeta, Shruthi G

Abstract: Agriculture is the best utility region especially inside the developing worldwide areas like India. Usage of records age in agriculture can substitute the circumstance of decision making and Farmers can yield in higher manner. About portion of the number of inhabitants in India relies upon on farming for its occupation however its commitment towards the GDP of India is just 14 percent. One suitable explanation behind this is the deficiency of adequate decision making by farmers on yield prediction. There isn't any framework in location to suggest farmer what plants to grow. The proposed machine learning approach aims at predicting the best yielded crop for a particular region by analyzing various atmospheric factors like rainfall, temperature, humidity etc., and land factors like soil pH, soil type including past records of crops grown. Finally our system is expected to predict the best yield based on dataset we have collected.

Index Terms: Crop yield prediction, Demand-based crops, Machine learning techniques, Random forest, Polynomial regression, Decision Tree, Supervised Learning

1. INTRODUCTION

Agriculture is something that individuals have started to finish up moderate on, disregarding that it's miles what is holding us alive. However, there is regardless some driving forward, enthusiastic ranchers whose life continues running on essentially developing. Regardless, there's in addition the pollution that is extending packages these days. The Main intention of the Department of Agricultural Marketing and Agricultural Business is to have a reasonable cost to the cultivating network who are pushed behind the current focused showcasing situation and the mission of accomplishing the reasonable cost is by making the current demonstration and principles solid and progressively compelling by executing new innovations and systems went for lessening pre and post-gather misfortunes through legitimate and sorted out techniques and urge enhancing the market. The vital motivation behind making a managed market is to put off the undesirable exchange work out, to diminish the charges inside the commercial centre and to offer reasonable expenses to the Farmers. A few activities have been taken to advance rural showcasing a decent method to cultivate and keep up the place of country monetary improvement. To advantage the cultivating from the new worldwide market get admission to potential outcomes, the inward rural promoting device inside the United States of America moreover wishes to be joined and strengthened. In interesting, the commercial centre contraption must be revived to:

- Provide impetuses to Farmer to deliver more.
- Pass on the changing over wishes of the purchasers to the makers to empower producing making arrangements.
- Foster genuine challenge a considerable lot of the market players and
- To improve the offer of Farmers in the last expense of his rural produce.

Today the farmers develop crops dependent on the experience picked up from the past age .Since the customary technique for cultivating is polished there exists an overabundance or shortage of yields without gathering the real necessity. The farmers don't know about the interest that happens in the current horticultural economy. This results in the misfortune to the Farmers. The communicated thought processes arranged by significance in the back of Farmer suicides have been condition, low produce costs, weight and hover of relative's obligations, poor water system, and blast inside the cost of development. The primary reason is the low costs of the items and the expanded expense of development. The expenses of yields are controlled by economic interest and the points of confinement of the creation. Yield forecast is one of the undertakings that should be possible by bleeding edge ML calculations. Field sensors, satellites, unmanned flying engines (UAVs), and cultivating hardware can give a goliath amount of records on soil circumstances, plant physiology, climate, and several of the procedures taking locale in a homestead. These datasets license the approach of sort and estimate molds that might be very useful to Agriculture generation. India is growing quick in populace. The call for is high and could blast in coming predetermination along these lines, to make certain sustenance security vertical advancement in farming is the need of great importance. For this a blended basic and methodological methodology likes assortment commencement, pesticide and composts the executives, consolidated editing, water collecting, proficient water system techniques and numerous others may be required. Additionally, it transforms into basic to re-enact and expect the harvest yield underneath encompassing circumstances past to the usage arrange for viable yield the executives and favored outcomes, more noteworthy so in the sprinkled locale and when India is inclining nearer to exactness cultivating researchers. Since the relatives between harvest yield and the climate and non-atmosphere factors are non-straight and comprise of fair dimension difficulties, machine examining may demonstrate a triumph elective for yield expectations.

Scope of the Project

The scope of the project is to determine the crop yield of an area by considering dataset with some features which are important or related to crop production such as temperature, moisture, rainfall, and production of the crop in previous years. To predict a continuous value, regression models are used. It is a supervised technique. The coefficients are preprocessed and fit into the trained data during training and construction the regression model. The main focus here is to reduce the cost function by finding the best fit-line. The output function facilitates in error measurement. During training period, error between the predicted and actual values is reduced in order to minimize error function. Python is used for this project by

using Jupyter platform for simulation.

Problem Statement
To Design, Develop and Implement the training model by using different inputs data. So machine will able to learn the features and extract the crop yield from the data by using machine leaning techniques.

Objectives of the project
The proposed system aims at predicting or forecasting the crop yield by learning the past data of the farming land. By considering various factors such as soil conditions, rainfall, temperature, yield and other entities the system builds a predicting a model using machine learning techniques. Here we make use of different machine learning techniques such random forest, Polynomial Regression, Decision Tree. Performance is evaluated based on predicted accuracy.

Challenges
Challenges are the major basis which imminent the negative impacts on current project. Some of the challenges faced during crop yield prediction are:

- Choosing appropriate dataset, after choosing dataset tuning of the parameters which makes project more efficient to get the desired results.
- Model must be trained by taking consideration of less computational efficiency and power.
- Increase of error rate due to dynamically changing the environment.

2 LITERATURE REVIEW

Detailed submission guidelines can be found on the author Grey prediction system which gives an excellent prediction accuracy of price forecast in production market, is made used in this work. This forecast structure is used to predict the market costs of different yields. By implementing demand grade for each crop, the real downside of this framework is destroyed i.e., price of the crops will not be stable all the times [1]. One of the useful support system in Bangladesh focuses on helping the poor farmers by assisting the demands about the crop through website. The drawback in the system is that the uneducated farmers were not able to use the system, even if the farmer knows about the system they could not able to operate the model. In this model the data will be sent by means of SMS voice message in regional language [2]. The procedure of crop yield prediction is done by using Data Mining approach which results in prediction of analyzed soil dataset. The interest existing in the rural economy is not considered by the system. This system overcomes the drawback by considering the demands based on the market price crops and it is suggested to the farmers for better growth [3]. To improvise the value and gain of farming area, data mining techniques are made used which selects the appropriate crops for cultivation and predicting the crop yield. Feasible suggestions to farmers and meeting the current demands are not provided by this system which serves as a drawback here [4]. A novel system known as extensible Crop Yield prediction framework is built for precision agriculture using data mining techniques. In this paper there is an investigation of requirement for crop yield prediction and different systems have been utilized and finally it results in a framework which is flexible for prediction accuracy [5]. In this work the author has used various data mining algorithms like Naïve Bayes and KNN to predict the class of analyzed soil dataset. The soil is categorized into high, medium and low. By doing this the farmer and the soil analyst gets the prior knowledge about the land. Meanwhile they can decide which crop best suits to sows. The results in-turn will help in predicting the crop yield [6]. Agriculture is one of the tedious processes which involve numerous estimations and effecting factors. In this research work author focused on modeling some of the important inputs which plays a major role in the collected dataset and to derive a strong relationship between the variables. SVM and k-means algorithms are used to forward pollution from atmosphere and also to classify between soil and plants [7]. Fuzzy Cognitive Maps are one of the Soft Computing techniques used for modeling expert's knowledge. In this work the author has made use of this approach to predict yield of cotton plants. Representing the knowledge more visually, effectively, simple and structural makes this technique more advantageous and thus a convenient approach for predicting and improving cotton yield prediction [8]. The author proposes a model which focuses on removing noise factors to get efficient predictions. By considering various factors like differences in agricultural policies and practices. To accommodate the spatial dependence arising between different regions prior distributions are developed. In addition to this basis expansions and dimension-reduction schemes are incorporated to evaluate the improved predictions [9]. To achieve excellence farming and suggest the farmer to select the best previous agriculture information is used. This research work also paves way to improvised yield prediction and increase in income level of crops. [10]

3 METHODOLOGY

Machine learning mainly consists of three learning methods, namely supervised learning, reinforcement learning and unsupervised learning methods of training a model. Supervised learning is a learning method which maps known input resulted into output which maps from input to output. But in case of unsupervised leaning we would not know targeted output in this learning we should train the model in order to get desired output.

3.1 Dataset Description

Generally researchers used .csv files of agriculture dataset for crop yield prediction. The dataset is supervised learning. It consists of different attributes like County Name, State, humidity, temperature, NDVI, wind Speed, yield etc.

3.2 Proposed Framework

Current paper focus on Machine learning which includes supervised learning models. System architecture is one which defines the Conceptual model in different structures and multiple views of the system. Fig. 1 shows architectural design of proposed system of the project.

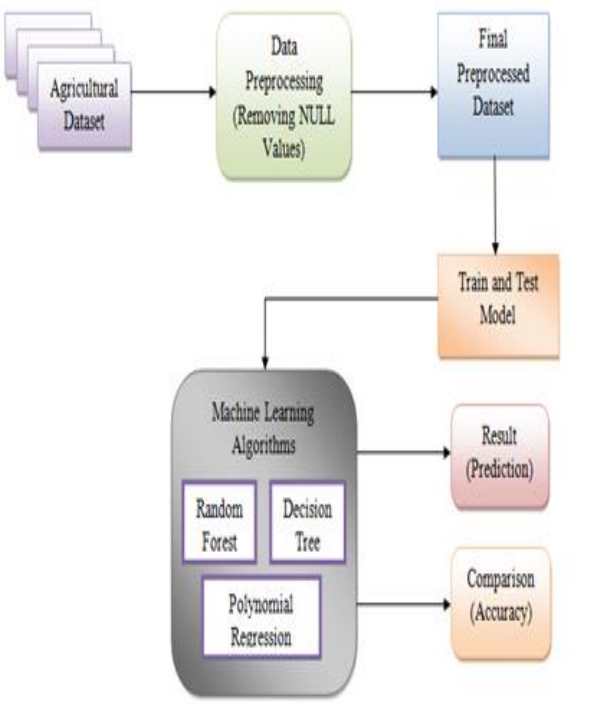

***Fig.1**. Block diagram of proposed system*

Above architecture clearly explains about how the components of the system communicate among themselves starting from preprocessing of data. This proposed framework is able to finding out the crop yield. This model gives clear picture of huge amount of data capture and preprocessing of data to remove the unwanted data such as NULL etc presented in it. During preprocessing step we spit the dataset into training and testing dataset. Train dataset to detect the crop yield present in the dataset using appropriate supervised learning algorithms. Apply the machine learning techniques which are helpful for finding crop yield for any of new data occurred in the data. After this data acquisition suitable machine learning algorithm must be applied to compute efficiency and capability of the model, here we have applied various machine learning algorithms like random forest, Polynomial Regression, Decision Tree etc. Metrics like accuracy, precision will be calculated for the proposed model. This system architecture focuses 3 parts such as flow data, Machine learning techniques, and modules for detecting crop yield and feature selection modules.

3.3 Algorithms

Polynomial regression

In this model regression analysis is done by deriving the relationship between x independent variable and y dependent variable as nth degree polynomial of x. A non-linear relationship is fitted to regression model x and mean y, which is denoted by E(y | x), and is used for phenomenon's like growth rate of tissues, sediments of carbon isotopes in lakes, and epidemics disease progression. Despite the fact that polynomial regression fits a non-linear model, as a statistical estimation it is considered as linear, as a function E(y | x) as unknown parameters from the data. Due to this polynomial regression is considered to be special case of multiple-linear regression.

Decision Tree

As decision tree employs greedy method, attribute chosen in the first step cannot be used later to give better classification of data. If at all it is used in the next steps Decision Tree over fits the training dataset that can lead to poor results. Ensemble model is incorporated to overcome this drawback and promising results are obtained by ensemble models.

Random Forests

Random forest, as the name says it is a combination of number of decision trees and an ensemble classification model. Random forest model collects trained data from all the tree nodes and separates the weaker nodes training data to get better predictions. Both classification and regression problems are solved using RF model.

4 RESULTS AND DISCUSSION

In this research, Machine learning frameworks played an important role in performing this experiment, this experiment is carried on personal notebook DELL, which has a configuration of Intel Core i3- 4300U CPU @1.90 GHz, and 8 GB of memory. Importance of the project results are examined in this section. In this project we have chosen 2 files of dataset which are in .csv format has a total size of 5l.9MB of data we have considered for the experiment. There will be 70% of data for training and 30% for testing. Initially we are choosing samples for training of model and chosen some of the features like Yield, temperature and Location etc. After successful training and the testing of dataset we moved further for finding the accuracy the model. The accuracy of the model shows that how we have predicted the yield of the crop in compare to original data the more the accuracy of the model is near to the original yield value. Data analysis is done by plotting the yield variable with different attributes like state, pressure, temperatureMax, temperatureMin etc.

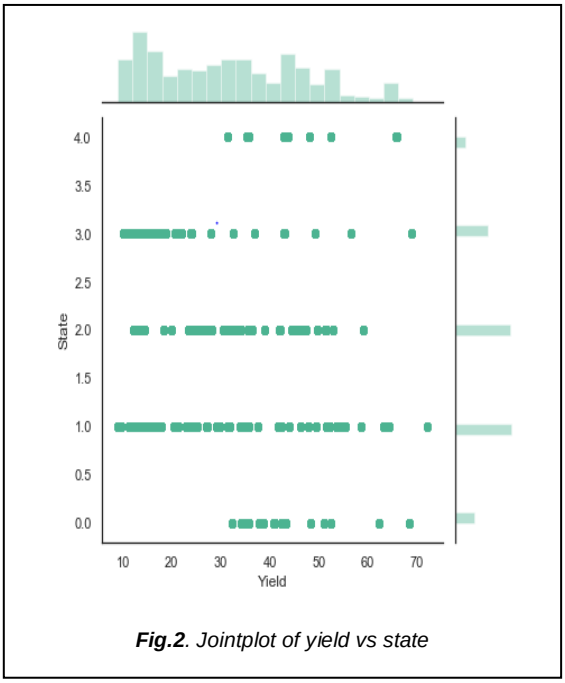

***Fig.2**. Jointplot of yield vs state*

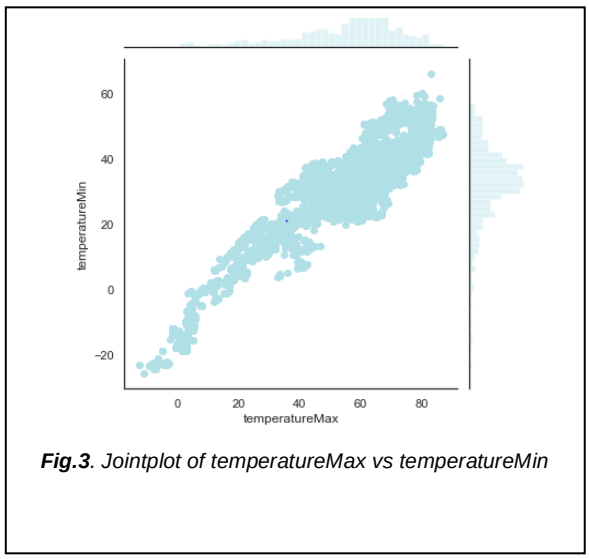

***Fig.3**. Jointplot of temperatureMax vs temperatureMin*

Accuracy of Polynomial Regression

The following figure shows the accuracy of Polynomial Regression Model. Where we have achieved testing accuracy of 88 percent.

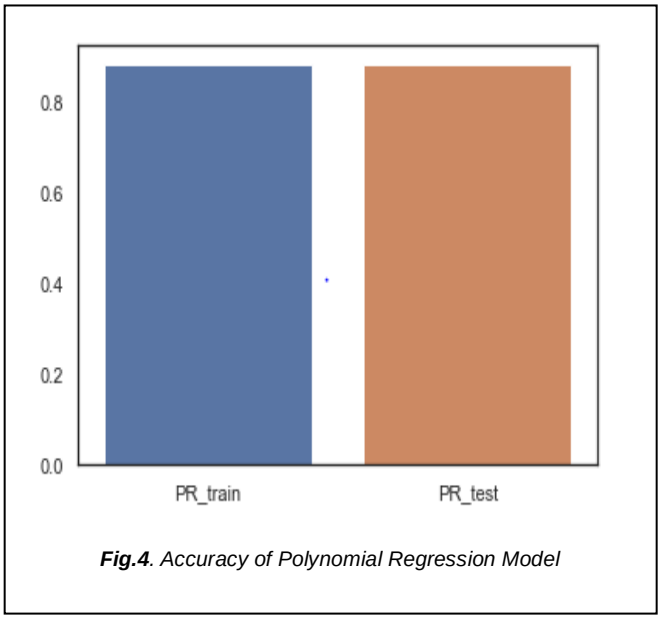

***Fig.4**. Accuracy of Polynomial Regression Model*

Accuracy of Decision Tree Regression

The following figure shows the accuracy of Decision Tree Regression Model. Where we have achieved testing accuracy of 74 percent.

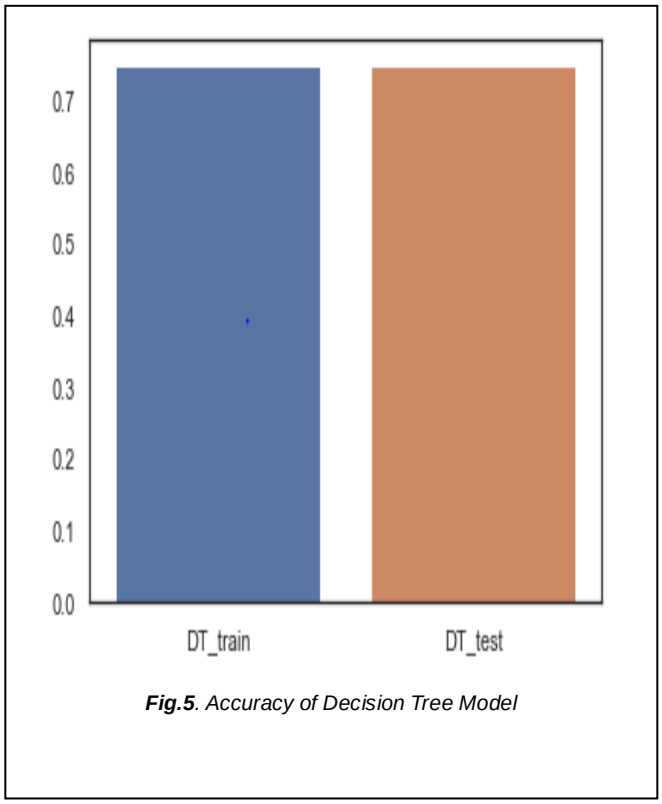

***Fig.5**. Accuracy of Decision Tree Model*

Accuracy of Random Forest Regression

The following figure shows the accuracy of the Random Forest Regression model. Where we have achieved training and testing accuracy of 97 and 85 percent respectively. This figure shows the accuracy with n_estimators = 10

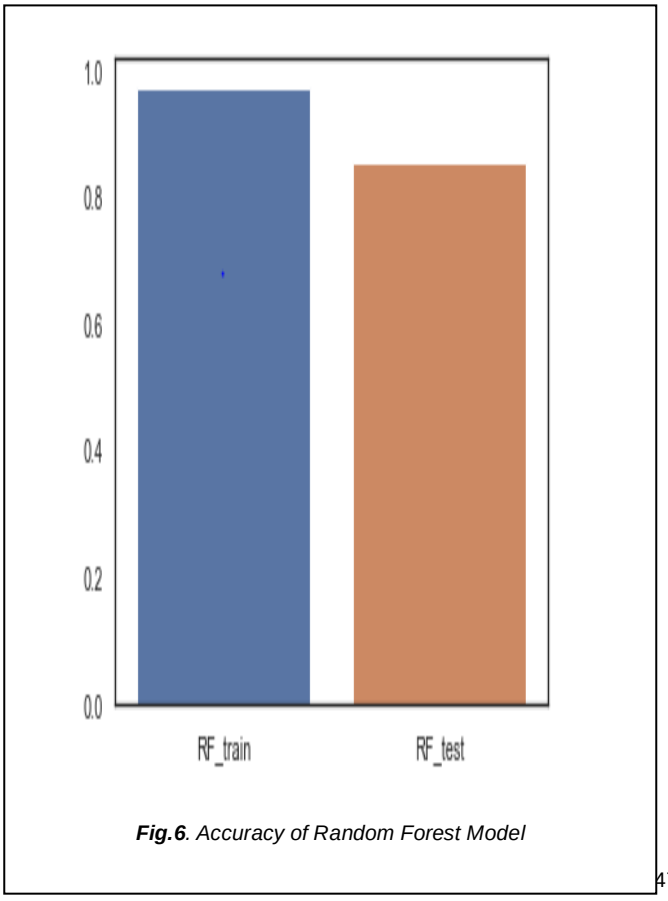

***Fig.6**. Accuracy of Random Forest Model*

Accuracy with n_estimators =100, we can see the increase in accuracy of training and testing of 98 and 88 percent respectively with increase in time complexity.

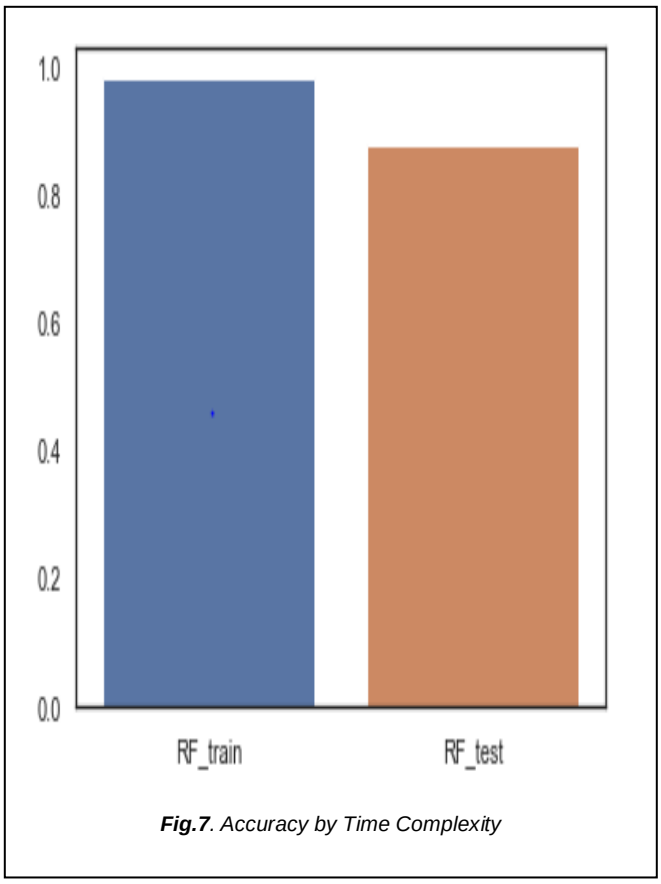

***Fig.7**. Accuracy by Time Complexity*

Further when the data is divided into different ratio its result in varied accuracy. When the training and testing data is split into 50% and 50% then accuracy of
Polynomial regression is 88% and
Decision Tree is 68% and
Random Forest is 97 and 88%

When the training and testing data is split into 75% and 25% then accuracy of
Polynomial regression is 88% and
Decision Tree is 77% and
Random Forest is 98 and 88%

5 CONCLUSION

This project is undertaken using machine learning and evaluates the performance by using Random forest, Polynomial Regression and Decision Tree algorithms. In our proposed model among all the three algorithm Random forest gives the better yield prediction as compared to other algorithms. Along with random forest, Polynomial Regression, Decision Tree model classify the output that shows improvements in dataset. So we analyzed that proposed model has got more efficiency than the existing model for finding crop yield. The implementation of above system would help in better cultivation of the agricultural practices of our country. Further it can be used to reduce the loss faced by the farmers and improve the crop yield to get better capital in agriculture. The model can be improved by integrating this with other departments like horticulture, sericulture, and others towards the agricultural development of our country.

ACKNOWLEDGMENT

I mainly thanks to my guide Prof. Shruthi G who has provided the extensive guidance and expertise throughout the study. Thanks to all who helped me in completing this work.

REFERENCES

[1] J.iajun Zong, Quanyin Zhu, “Apply Grey Prediction in the Agriculture Production Price”, Fourth International Conference on Multimedia Information Networking and Security, 2012.
[2] Michael gurstein, “a decision support system to assist the rural poor in Bangladesh”, IEEE TECHNOLOGY AND SOCIETY MAGAZINES, September 2013.
[3] Monali Paul, Santhosh K. Vishwakarma, Ashok Verma, “Prediction of Crop Yield using Data Mining Approach” Computational Intelligence and Communication Networks(CICN), International Conference 12-14 Dec. 2015.
[4] Tng Zhang, “Solving large scale linear prediction algorithm”, proceedings of the twenty-first international conference on Machine Learning Aakunuri Manjula, G. Narsimha, “XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction”, IEEE Sponsored 9th ISCO, 2015.
[5] Aakunuri Manjula, G. Narsimha, “XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction”, IEEE Sponsored 9th ISCO, 2015.
[6] D. Ramesh, B. Vishnu Vardhan, “Analysis Of Crop Yield Prediction Using Data Mining Techniques”, 2015
[7] A.A. Raorane, R. V. Kulkarni, “Data Mining: An effective tool for yield estimation in the agricultural sector”< IJETTCS, vol. 1, no. 2, pp. 75-79, 2012.
[8] E. I. Papageorgioua, A. T. Markinosb, T. A. Gemtosb, “Fuzzy cognitive map based approach for predicting yield on cotton crop production as a basis for decision support system in predtiction agriculture application”, Elsevier, vol. 0, no. 11, pp. 3643-3657, 2011
[9] Luke Bornn, James V. Zidek, “Efficient stabilization of crop yield prediction in the Canadian Prairies”, Elsevier, vol. 0, no. 0, pp. 223-232, 2012.
[10] R.Sujatha and P.Isakki, “Crop Yield estimation using classification techniques” Computational Intelligence and communication Network (CICN), International Conference 7-9 Jan. 2016.
[11] Rakesh Kumar, M.P. Singh, Prabhat Kumar And J.P. Singh, “Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique” International Conference 6-8 May. 2015.
[12] E. Manjula, S. Djodiltachoumy, “A Model for Prediction of Crop Yield” International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.
[13] S. Nagini, Dr. T. V. Rajini Kanth, B.V.Kiranmayee, “Agriculture Yield Prediction Using Prediction Analytic Techniques” International Conference on Contemporary Computing and Informatics 2016.
[14] Yung-Hsing peng, Chin-Shun Hsu, and Po-Chuang Huang, “Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets” Nov. 20-22, 2015.
[15] S. Veenadhari, Dr Bharat Misra, Dr. CD Singh, “Machine Learning approach for forecasting crop yield based on climatic parameters” 2014 International Conference on Computer

Communication and Informatics (ICCCI -2014), Jan. 03-05, 2014, Coimbatore, INDIA.