ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 10, October 2016**

# Annual Crop Yield Prediction and Recommend Planting of Different Crops by Using Data Mining Technique

Naushina Farheen Imam Shaikh[1], Prof. R. V. Argiddi[2]

PG Student, Department of Computer Science and Engineering WIT, Solapur, Maharashtra, India[1]

Asst. Professor, Department of Computer Science and Engineering WIT, Solapur, Maharashtra, India[2]

**ABSTRACT**: In our nation, farmers are not getting expected crop yield from their productions. Crop production mostly depends on weather conditions and some statistical methodologies. To get higher crop production yield, farmers sometimes need advices for predicting and analysing future crop production. This helps farmers to produce a crop with maximum yield. Such methods will be helpful for farmers and government to make a better decision to increase crop production. Our main focus in this paper is on data mining. Data mining is a technology which extracts the data and gets some new pattern from it. We proposed to implement K-NN technique for creating clusters and making predictions from vast amount of data. And we will use geospatial analysis for crop yield prediction. GSA method will be applied to the extracted data to identify patterns in the field. Once patterns and correlations are discovered, previous knowledge or patterns will be modified to optimize yield and production costs, and minimize environmental impact.

## I. INTRODUCTION

Today, India is at the second position in world for farm output. As with increasing economic sector, the economic contribution of farm is getting decreased. As we know agriculture is main economic sector of India which plays an important role for economic growth. India is the world's largest producer of many crops like fresh fruits and vegetables, milk, fabric crops, several other crops such as castoroil seeds. India is at second position in production of wheat and rice.

In India, most of the farmers are not getting the expected crop yield due to several reasons. The agricultural crop yield is primarily depends on weather conditions. The farmers necessarily require continuous advice to predict the production of future crop and analysis must be made which helps the farmers to maximize their crop production. Investigation of how to improve the productivity and economic growth of agricultural production is one strategy which is based on with challenges of climate change and available agricultural land. Most of research and trials have been undertaken to analyze land, soils, and climate.

Yield prediction is an important agricultural problem. Each and Every farmer is always tries to know, how much yield he will get from his expectation. In the past, yield prediction was calculated by analysing farmer's previous experience on a particular crop. The Agricultural yield is primarily depends on weather conditions, pests and planning of harvest operation. Accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management.

Historical information about crop production yield is also important for supply chain operation of companies who engaged in industries that use agricultural production as raw material. Historical crop production information like crop size and related risks about crop helps these companies to take supply chain decision like scheduling the crop production [9].

Data mining is an important research field for agriculture. Some tools Data mining are more powerful for generating rules from large amount of data. We can use such tools for generating patterns or knowledge from large amount of agriculture dataset. Generally, data mining is a technology which extracts the data and summarizes it into useful and accurate information.

We proposed to implement one of the data mining tools for analysing, extracting and predicting the agricultural information. We proposed to use k-nearest neighbour technique which creates the clusters predict the data sets and provide the summarized and accurate patterns from original data. We proposed to implement fuzzy logic for optimizing

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 10, October 2016**

the extracted solution. And finally, we will use geospatial analysis for crop yield prediction. Geospatial analysts filter out relevant and accurate information and apply it to conceptualize and visualize the order hidden within the apparent disorder of geographically sorted data. This will allows them to provide accurate trend analysis, making patterns and predictions. Once patterns and correlations are discovered, previous crop production solutions will be modified to optimize yield and production costs, and minimize environmental impact.

## II. LITERATURESURVEY

Sam Y. Sung, Member, IEEE Computer Society, Zhao Li, Chew L. Tan, and Peter A. Ng was implemented "Forecasting Association Rules Using Existing Data Sets".In this paper, they implemented a model for understanding the differences between two situations makes impact on the changes of the rule. Using this model, the system was provided a Combination Data Set technique, to get a good estimation of the set of rules for a new situation. [1]

YETHIRAJ N G was focused on "Applying Data Mining Techniques In the field Of Agriculture And Allied Sciences". In this paper, the author made the review on data mining tools for predicting agricultural data. Some of the techniques, such asID3 algorithms, the k-means, the k nearest neighbour, artificial neural networks and support vector machines applied for agriculture. Data mining in agriculture is a new approach for predicting agricultural crop production. [2]

S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh was made a review on "Data mining Techniques for Predicting Crop Productivity". In this paper, review was made on the research of applications of data mining techniques in the field of agriculture. In review, they provided methodology information, advantages, disadvantages and other information related to each technique. [3]

Ramesh A. Medar , Vijay .S. Rajpurohit was worked on " A survey on Data Mining Techniques for Crop Yield Prediction".In this paper, they mainly focused and implemented the applications of data mining techniques in agricultural field. Number of Data Mining techniques such as K-Means, K-Nearest Neighbour (KNN), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for very recent applications of data mining techniques was used for crop prediction output [4]

DildarKhan T. Pathan1, Pushkar D. Joshi2, Prof. S. U. Balvir were implemented "Prediction of soil quality for agriculture" In this paper they provided web based application for the soil testing as well as free messages for the farmer which contains information like soil testing code, fertilizer information for the crop and also provided crop production advice. The reports based on soil were generated. [5]

Raorane A.A. , Kulkarni R.V. were implemented "Data Mining: An effective tool for yield estimation in the agricultural sector". here they figure out the accurate information about historical yield of crop and important modelling input, which helps farmers & Government for taking decisions for using proper policies. They had implemented new methods and techniques such as data mining that can extract the knowledge of the data to the crop yield judgment. [6]

Jeysenthil.KMS, Manikandan.T, Murali.E implemented "Third Generation Agricultural Support System Development Using Data mining". They had implemented various classification techniques of data mining and apply them to a soil. In data mining conception, clustering and classification technique produced better solution to the farmers about their cultivation. [7]

Mohammad MotiurRahman, NaheenaHaq and Rashedur M Rahman "Comparative Study of Forecasting Models on Clustered Region of Bangladesh to Predict Rice Yield" In this study a research initiative has been taken to predict the yield of crops using machine learning models. The models were at first trained on the correlation between past environmental patterns and crop production rate. Then the models are compared to measure their effectiveness on unknown climatic variables [8].

Chapman P. Gleason "Large area yield estimation/forecasting using plant process model" This paper describes the history and purpose of the Statistical Reporting Service,current operational yield forecasting and estimation

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

methodology, and potential application of plant process models (PPM) to large area yield forecasting. Potential problems, statistical research and the future of PPM for forecasting large area yields is also discussed.[9]

## III. PROPOSED METHODOLOGY

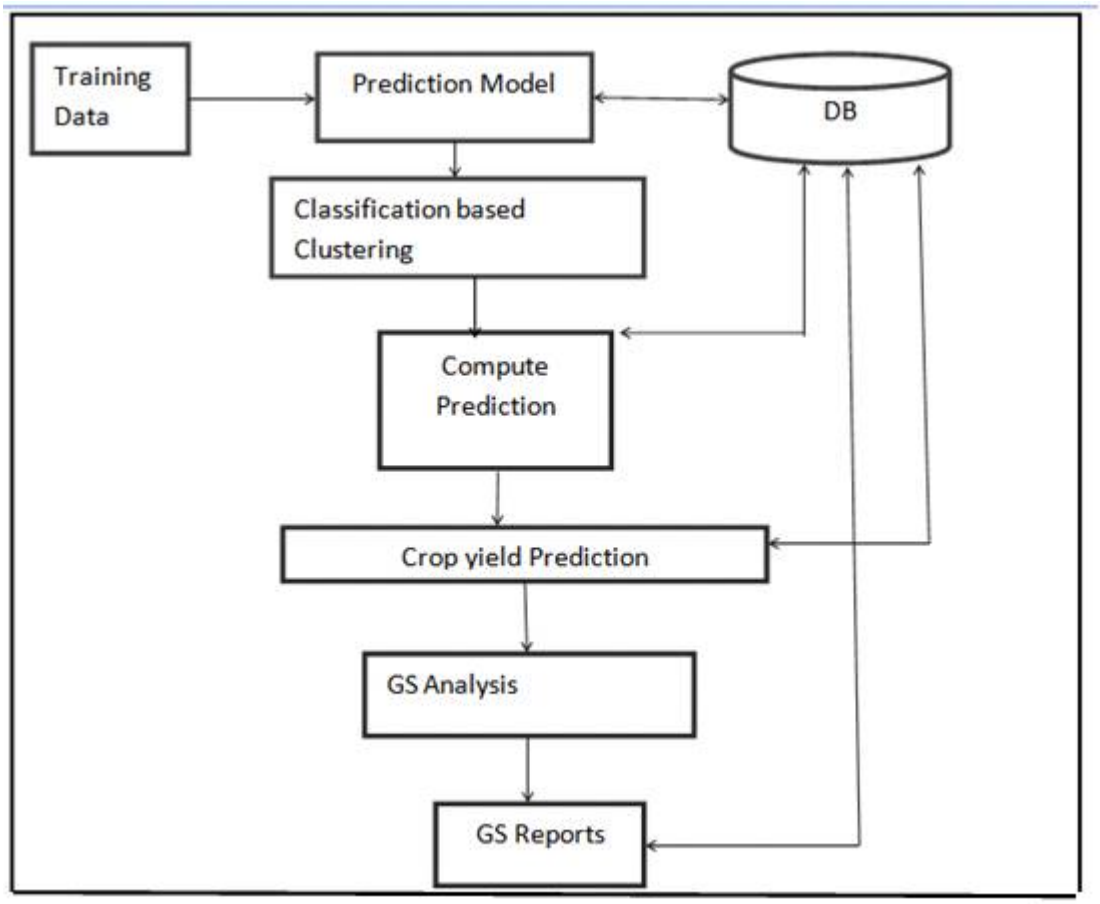


Figure 1: System Architecture

Fig.1 shows the architecture of crop prediction which includes an input module which is based on agriculture land and environment. The crop knowledge is consisting of farm knowledge. The agricultural data sets are clustered based on similar contents. Then prediction rules will be applied to output of clustering to get results in terms of crop and cost then geospatial analysis is done and finally we get geospatial result.

**K-Nearest neighbour technology:**
K-NN technology gives the output as a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its *k* nearest neighbours.

**Fuzzy Logic:**
Fuzzy logic is a mathematical approach used for computing and optimizing data based on “degrees of truth" instead of giving true or false values.

**Geospatial analysis:**
Geospatial analysis is an approach for applying statistical analysis and other analytic techniques to data which has a geographical or spatial aspect.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 10, October 2016**

**Training data:**
A **training** data set is a set of data used to discover predictive relationships. **Training** set is used in intelligent systems, machine learning, genetic programming and statistics.

**Testing data:**
**Test data** is the input given to a software program. It represents data that affects or is affected by the execution of the specific module.

**Our proposed system works as follows:**

**Step 1: Giving input (training data and testing data)**
We provide input to system as training data and testing data.

**Step 2: compute the prediction:**
Testing data and training data is given to K-NN algorithm which creates clusters from given data. K-NN algorithm gives relevant data, patterns from given datasets.

**Step 3: Optimize the predictions:**
Once we get the relevant clusters we will use fuzzy logic, which creates the membership functions. It optimizes the extracted clusters and provide predictive model.

**Step 4: GSA reports are generated**
The optimized clusters are given to the GSA analysis model. The GSA analysis filters out relevant information and generates the reports which contain the advices for crop productions for rising economics, historical information and fertilizers information.

## IV. CONCLUSION

Now a day, Farmers are not getting expected crop production yield. For getting expected crop yield, farmers need advices for predicting and analysing future crop production, so that farmers can easily take a decision before any crop production. This helps farmers to higher crop yield. In this paper, we proposed to implement K-NN technique for creating clusters and making predictions from vast amount of data. And we will use geospatial analysis for crop yield prediction and GSA reports which contains advices for farmers, fertilizers information.

## REFERENCES

[1] Sam Y. Sung, Member, IEEE Computer Society, Zhao Li,Chew L. Tan, and Peter A. Ng, "Forecasting Association Rules Using Existing Data Sets", IEEE transactions on knowledge and data engineering, vol 15, no.6, November/December 2003.
[2] YETHIRAJ N G. "APPLYING DATA MINING TECHNIQUES IN THE FIELD OF AGRICULTURE AND ALLIED SCIENCES" International Journal of Business IntelligentVol 01, Issue 02, December 2012
[3] Data mining Techniques for Predicting Crop Productivity – A review article S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh IJCST Vol. 2, Issue 1, March 2011
[4] 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, September 2014
[5] DildarKhan T. Pathan1, Pushkar D. Joshi2, Prof. S. U. Balvir "Prediction of soil quality for agriculture" International Research Journal Of Sustainable Science & Engineering IRJSSE / Volume :2 / Issue: 3 / March. 2014
[6] Raorane A.A. , Kulkarni R.V. "Data Mining: An effective tooforyieldestimation in the agricultural sector"International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 1, Issue 2, July – August 2012
[7] Jeysenthil.KMS1, Manikandan.T2, Murali.E3 "Third Generation Agricultural Support System Development Using Data Mining" International Journal of Innovative Research in Science, Engineering and Technology Vol. 3, Issue 3, March 2014
[8] Mohammad MotiurRahman, NaheenaHaq and Rashedur M Rahman "Comparative Study of Forecasting Models on Clustered Region of Bangladesh to Predict Rice Yield", 17th.IEEE International Conference on Computer and Information Technology (ICCIT), Dhaka, 2014.
[9]Chapman P. Gleason LARGE AREA YIELD ESTIMATION/ FORECASTING USING PLANTPROCESS MODELS By Chapman P. Gleason For Presentation at the 1982 Winter Meeting AMERICANSOCIETY OF AGRICULTURAL ENGINEERS Palmer House, Chicago, Illinois December r 14-17, 1982.