International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

# RICE YIELD PREDICTION USING DATA MINING TECHNIQUE

**Nanjesh Gowda M P[1], Chetan[2], Ramya S [3]**

*[1, 2]BE, Department of CSE, NIE Mysore, Karnataka, India*
*[3]Assistant Professor, Department of CSE, NIE Mysore, Karnataka, India*

---

**Abstract -** *- Food is the basic need for all the human beings in the world without food it is impossible to the human beings to survive. Storing of crops also plays a major role for all the countries especially for the developing one. In the previous days farmers have to take the sample of soil from their region then check the nutrient of the soil then only they can grow crop. It is a time consuming process for the farmers and if the weather changes there will be loss on that crop. To overcome this problems in our paper we take the soil from the farmers from their respective region and predict the amount of rice can be grown in that region by considering the various parameters such as pH, EC, N, P, K etc from this the farmer can know how much of rice can be grown in his region. Here we are considering a large number of dataset to predict the yield of the rice. In this paper we are using various data mining technique to predict the yield of the rice. By using the KNN algorithm it gives the accurate result to the farmers.*

***Key Words*: Data Mining, Yield Prediction, K- Nearest Neighbour (KNN), Database.**

## 1. INTRODUCTION

Karnataka is a state where most of the people in the district their main occupation is agriculture. India is a state where agriculture is the backbone of the country. It is the major source of economy for their families. India is a country which has many seasons. The proper planning and proper sowing only give the expected yield otherwise there is a loss to the farmers. The main aspect of the governments of such countries is to fulfill enough storing of crop for long term, mainley at a time of natural disaster. In this paper, we attempt to forecast yield of rice with astute analysis.

The Department of Agriculture has been created many programmers to educate the farmers to grow proper crop in proper time. The government has introduced many classes to the farmers to transfer the latest technical knowledge to farming, from this their yield can also increased in production and productivity. Since rice cannot be grown in all the regions of the state, to grow the crop rice the region must support to the climate, humidity, rainfall, soil features…If any of the above is not support their will be loss to the farmers. Hence taking necessary measures to gain maximum yield of rice during the respective seasons become the main priority of the farmers. Storing huge amount of data has given us to find the patterns and create models from this we can predict the yield of the rice accurately in the future. The main objective of this research is to predict the yield of the rice.

### 1.1 Prediction System

In our paper we focus on Data Mining technique to predict Rice yield though different parameters. The rice prediction is done by Data Mining using K-nearest neighbor (KNN) algorithm. The algorithm considers the parameters such as pH, EC, soil quality, Nitrogen, Potassium, Humidity, Rainfall…

### 1.2 K-nearest neighbor (KNN)

KNN is a simple algorithm that uses entire dataset during its training phase whenever prediction is required for unseen data. It searches through entire training dataset for k- most similar instances and data with most similar instance are returned.

#### 1.2.1 Features of K-nearest neighbor's

- KNN stores the whole coaching dataset that it uses as its illustration.
- The prediction is fast in terms of time by calculating the similarities between the input sample data and each training data.
- It works on similarity measures.

## 2. RELATED WORK

Analysis of soil behavior and prediction of crop yield using data mining [1] in this paper the author has predicted the yield of the crops by using the Naive Bayes method. Here he takes the soil from the farmers check the soil parameters and weather conditions and predicted the yield of the crop. The results of the crop are accurate. The problem of predicting the crop yield is formalized as a classification rule, where Naive Bayes and are used. Here the problem is it is suitable for crop yield but not suitable for rice crop type. It takes more time for prediction.

Navid Tanzeem Mahmood examines the nature, biology, geography and the soil parameters to predict the yield of the

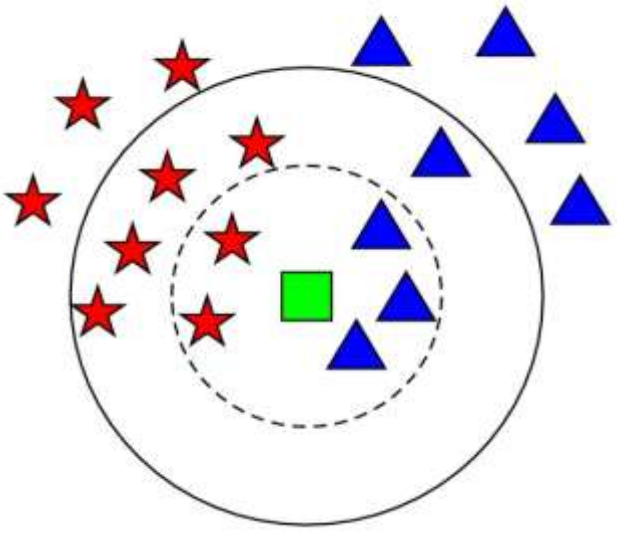
crop. In this paper several crop related data are given to the farmers to yield the maximum crop. Various methodologies are used to predict the yield of the various crops. In this paper data mining technique is used and linear regression is used. The use of the parameters are also minimum in the paper, the system predicts the output based on the temperature, rainfall, humidity, on the paddy growth does not support for the rice yield. The system uses the data mining technique hence it requires large number of data set [2]

Based on the above research papers, we are proposing a system which predicts the yield of the rice based on the given textual input by the staff. The machine which is trained using KNN (K- Nearest Neighbors) gives the output to the staff in the form of the tons and its related rice to be grown. The KNN is an algorithm used in supervised learning which stores all available data and classifies new data based on similarity measures.

## 3. METHODOLOGY

***K-nearest neighbor's classifier (KNN)***

K-Nearest neighbors (KNN) [3] is classification algorithm is a type of instant based learning or non generalizing learning. It does not attempt to construct a general internal problem but it compares the new problem with the training data that is stored in memory. The purpose of the k-nearest neighbour is to compares the data in the dataset and classify the new data points based on the similarities measure. Figure below shows nearest neighbours classification. KNN classification has two stages



1) Find the *k* instances in the dataset that is closest to instance *S*

2) These *k* number of instances then vote to determine the class of instance *S*

The Accuracy of KNN classifier depends on distance and K value. There are many ways of measuring the distance between two instances they are cosine, Euclidian distance. To evaluate the new unknown data, KNN computes its K nearest neighbours.

**Proposed Approach**

Our proposed system aims to enhance the performance of KNN classifier for rice yield predictions. Algorithm for our proposed system is shown below

Step 1: Input: Previous year's agriculture data which includes temperature, rainfall, humidity and other constraint..

Step 2: Output: Rice yield predicted

Step 3: Give the data set as input.

Step 4: perform KNN algorithm for input parameters.

Step 5: Euclidean distance between existing parameters and newly entered is calculated.

Step 6: Based on similarity measures output is predicted.

The nearest neighbor algorithm measures the distance $d_E(X_i, X_j)$ between query points $X_i$ and a set of training samples $X_j$ to classify a new object based on majority of K-nearest neighbor category of $Y$ attributes of training samples.

Query point $X_i = x_1, x_2, x_3, \ldots\ldots\ldots, x_n$

Training Sample $X_j = x_1, x_2, x_3, \ldots\ldots\ldots, x_n$

$$Dist(c_1, c_2) = \sqrt{\sum_{i=1}^{N} \left(attr_i(c_1) - attr_i(c_2)\right)^2}$$

$$k - NearestNeighbors = \{k - MIN(Dist(c_i, c_{test}))\}$$

$$prediction_{test} = \frac{1}{k}\sum_{i=1}^{k} class_i \; (or \; \frac{1}{k}\sum_{i=1}^{k} value_i)$$

## 4. FUTURE ENHANMENTS

The future enhancement of the project is to focus on not only rice but also various crops like maize, ragi, wheat…. Here we use KNN for prediction. Other algorithms like ID3 or C4.5 or Naive Bayes Algorithms can also be used. Future, we may add more parameters to the dataset which makes it more approachable to the farmers.

## 5. CONCLUSIONS

The study gives that the soil parameters and weather plays an important role in predicting the rice yield. From the dataset one can predict the yield but it is not accurate all the time if the there is change in weather there will be loss of yield. In this project we have use Data Mining to help farmers to increase their yield. It also provides the accurate prediction of the rice yield by considering the various parameters this avoids the loss of yield and economy to the farmers.

## ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of the any task would be incomplete without the mention of people who ceaseless co-operation made it possible whose constant guidance and encouragement crown all efforts. It gives us great pleasure to thank our guide Ms. Ramya S for their valuable suggestion and guidance during the process of the project and for having permitted us to work on this subject.

## REFERENCES

[1] Monali Paul, Santosh K. Vishwakarma, "Analysis of soil behavior and prediction of crop yield using data mining approach" IEEE 2015

[2] Navid Tanzeem Mahmood, Nazmul Hossain, "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh" IEEE 2015.

[3] D Ramesh, B Vishnu Vardhan. "Data Mining Techniques and Applications to Agricultural Yield Data" International Journal of Advance Computer and Communication Engineering Vol2, Issue9, September 2013.