

Paddy Yield Prediction Model Using Data Mining Techniques

Insha Sirur*, Karthik B, Sharath P, Mohan Kumari M, Rumana Anjum

Department of Computer Science, Vidya Vikas Institute of Engineering and Technology Mysuru

DOI: <https://doi.org/10.21467/proceedings.1.71>

* Corresponding author email: inshasirur024@gmail.com

Abstract

Prediction of agriculture yield is a job that requires unification of knowledge from several areas such as data mining, statistics and agriculture. Subject of crop Yield prediction has been very popular among various organizations working in agriculture, producers, etc. Prediction of crop yield helps in managing the storage of crops as well as it directs the transportation decisions, and risk management issues related to crops. Data mining focuses upon methodologies for extracting useful knowledge from data and there are several tools to extract the knowledge that is it is a proficiency of examining the dataset such that the end results can be deduced easily and rapidly from the dataset. Knowledge gathered can be used to forecast the paddy yield. We collected the data from different government organizations, after preprocessing of data applied k nearest neighbor algorithm using Data Mining using soil nutrients, fertilizers nutrients, rainfall and temperature.

Index Terms- data mining; k nearest neighbor; Yield prediction.

1 INTRODUCTION

Rice is one of the crop which is grown majorly in India as well as Asian countries. In India most of the farmers are not getting expected yield due to several reasons. Although crop Production has been coupled to a number of parameters such as evaporation, radiation, temperature, soil moisture and crop management practices. The objective of this research is to predict the future yield using temperature, rainfall, nutrients like nitrogen, potassium and phosphorous which are require for the soil and fertilizers and yield of previous year. Using k nearest neighbor algorithm yield will be calculated. Using data mining techniques many authors has been working- on agriculture. In [1] classification and clustering algorithm used to predict the yield. To predict the yield they were considered Rainfall, Area of sowing. In [2] another research authors used rainfall, temperature and rice yield to predict the future yield. For the yield prediction regression algorithm used. . In this research [3] presents a review on the use of artificial neural network in predicting crop yield using various crop performance factors. ANN can give better prediction compared to regression models. But it has disadvantage that too many nodes will becomes overstrained, causing it to memorize the training data resulting poor predictions and consumes lot of memory. In another research [4] author used k-means



© 2018 Copyright held by the author(s). Published by AIJR Publisher in Proceedings of the 3rd National Conference on Image Processing, Computing, Communication, Networking and Data Analytics (NCICCNDA 2018), April 28, 2018. This is an open access article under [Creative Commons Attribution-NonCommercial 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/) (CC BY-NC 4.0) license, which permits any non-commercial use, distribution, adaptation, and reproduction in any medium, as long as the original work is properly cited. ISBN: 978-81-936820-0-5

algorithm and various classifier for better accuracy. But it has disadvantage that if K value is too high, number of cluster will be more. And also it effects accuracy of results.

2 MATERIALS AND METHODS

2.1 The Study Area

Paddy is grown under various conditions. It is a plant which requires high heat and high humidity. It requires temperature between 20° to 30°C and annual rainfall is between 100cm to 200cm. Rice yield per hectare is low. So high yielding varieties of seed, fertilizers, insecticides etc should be introduced to increase rice yield per hectare. Loamy soil is most suitable for rice cultivation. Best production occurs in the alluvial soil of the river valley.

2.2 Data and sources

Data had been collected from the government sectors. Some of link which we got the data for yield prediction model those are agroman.in, des.kar.nic.in, agropedia.iitk.ac.in, www.yourarticlelibrary.com, <http://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics>, www.ricepedia.org, indianfertilizer.com, www.agritech.tnau.ac.in www.cropnutrition.com and <http://www.ksndmc.org/weather-info>.

2.3 Methodology

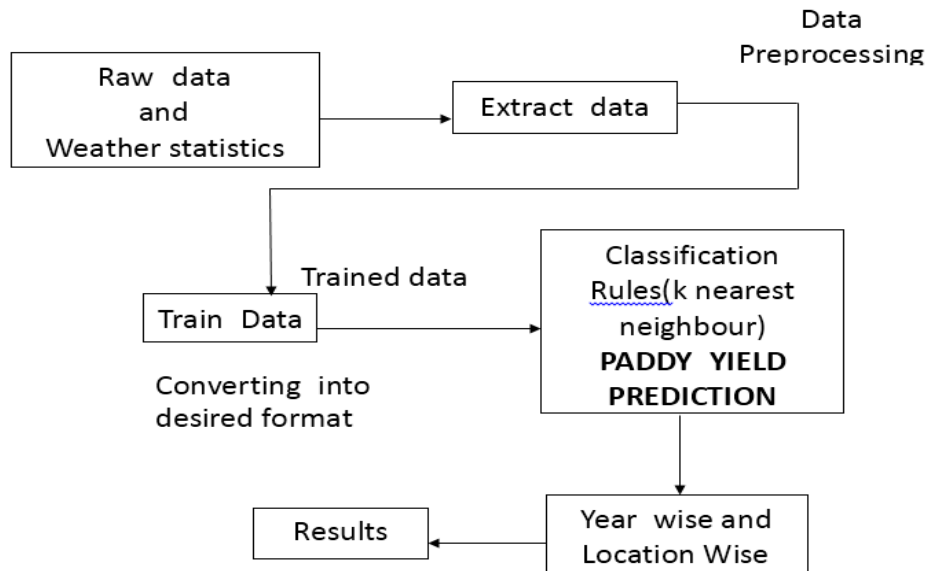


Fig 2.3.1 Methodology

The above figure shows the methodology. Weather information and Raw data like soil Nutrients, rainfall, temperature were collected from the government sectors. It preprocessed to remove unwanted data. After preprocessing it has converted into desired format and it give as input to the algorithm k nearest neighbor. And the system will predict the result in year wise and location wise.

2.4 Preprocessing

Real world data is often unfinished, uneven and/or lacking in certain trends and behaviors and is liable to contain many errors. Data preprocessing is a procedure of resolving such issues. It organizes raw data for further processing.

Data goes through the following steps in preprocessing component:

1) Data Cleaning: Data Cleaned through procedures such as filling missing values, ignoring extreme values, resolving inconsistencies.

2) Data Transformation: The data is aggregate and placed in an excel sheet.

3) Data Discretization: In this step the input and output attributes are discretization as low, high and medium based on some criteria like taking the mean values as medium and higher and lower values as high and low respectively.

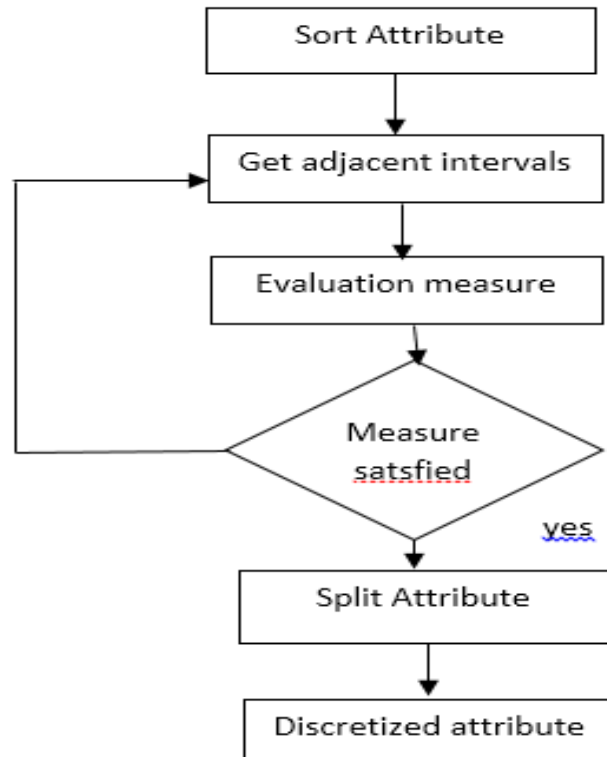


Fig 2.4.1 Discretization.

2.5 Algorithm

KNN algorithm is a simple algorithm that stores the previous values and classify future values based on similarity measure. KNN algorithm is easy to understand and implement.

Steps involved in working of k nearest neighbor

1. Determine K (number of nearest neighbor).
2. Calculate distance.
3. Determine K-Minimum distance neighbor.
4. Gather category Y values of nearest neighbor.
5. Use k nearest neighbors to predict value of query instance.

3 CONCLUSIONS

This model proves that weather conditions and nutrition like nitrogen, potassium and phosphorous play major role to predict the paddy yield. Using large data KNN algorithm predict accurate result. It helps farmers, future preparation in trade and loss minimizations. This model can enhance by considering level of osmosis, pesticides, wind blow and humidity.

References

- [1] Umid Kumar Dey, Abdullah Hasan Masud, "Rice Yield Prediction Model Using Data Mining" International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox's Bazar, Bangladesh.
- [2] Kuljith kaur, kanwalpreet singh attwal, "Effect of Temperature And Rainfall on Paddy Yield Using Data Mining" 7th International Conference on Cloud Computing, Data Science & Engineering 2017.