

Enhancing Crop Yield Forecasting in Tamil Nadu Through Data-Driven Approaches: A Comparative Study of Decision Tree and Random Forest Models

Jasmine M,

Assistant Professor,

Department of Computer Science with Artificial Intelligence,
Anna Adarsh College for Women

ABSTRACT

India is an agriculture-based country where more than 50% of the population is dependent on agriculture. Agriculture is a crucial aspect of India's economy. This thesis aims to provide a prediction model for the production of different crops in various districts of Tamil Nadu, taking into account the data of previous years. Rainfall and micro soil nutrients like Zinc(Zn), Iron(Fe), Copper(Cu), Manganese(Mn), Boron(B), Sulfur (S) are taken as major contributing factors for the prediction of crop production. This study utilizes data sourced from the Government of India's official open data platform, Data.gov.in. The findings of this study, including both analytical and predictive results, are presented. The prediction model was tested for the year 2016 and its result compared with the actual data for 2016. It was found that the model generated a crop yield prediction with a significant degree of accuracy. This kind of district-wise classification of crop yield and prediction of quantity of crop production in Tamil Nadu is a first of its kind in the literature, and seems a promising start to yield prediction for Tamil Nadu as also other states of the country.

Keywords: Crop Yield Prediction, Tamil Nadu Agriculture, Soil Nutrients and Rainfall, District-wise Crop Analysis, Data-Driven Agriculture

INTRODUCTION

India is predominantly an agriculture-based economy. In earlier days people cultivated the crops in their own land and were able to take care of their own needs. But with population steadily increasing, and demands consequently increasing as well, it has become the need of the hour to increase crop production and use technology on a mass scale to ensure highest possible crop yield.

An important problem in agriculture is to know how much would be the yield of a crop in a given piece of land in the near future. Such prediction about the possible yield in his land could be of potential use to the farmer, to decide on the type of crop that he should sow in his land in the next cycle. In the past, yield prediction was just guess work by the farmer, by his analyzing his experience with a particular crop in his piece of land. However, technology could be used effectively in this area, to

predict the crop yield levels of various crops, given the soil nutrient levels and rainfall. This work seeks to do precisely this. This work seeks to predict the yield of different crops in 31 districts of Tamil Nadu by considering the micro soil nutrient levels in the area and the past rainfall patterns.

LITERATURE SURVEY

This project builds a soil fertility testing model that recommends suitable crops based on sensor data. It also suggests fertilizers improve productivity and shows region-wise crop data through graphs. Farmers can register, chat with experts, and share ideas. The app helps analyze soil health, choose profitable crops, and locate nearby fertilizer shops.

Vijayabaskar et al [2] has proposed a data-driven approach to crop prediction using predictive analytics techniques. By leveraging historical agricultural data, the research focuses on identifying patterns and trends that influence crop yield. The model incorporates various parameters such as soil type, weather conditions, and crop characteristics to enhance prediction accuracy. Decision trees and classification algorithms are used to analyze the data and predict the most suitable crop for a given set of environmental and soil conditions. The paper demonstrates the potential of predictive analytics in assisting farmers and agricultural planners to make informed decisions, ultimately improving productivity and resource utilization. This work highlights the significance of integrating data mining techniques in the agricultural domain to tackle challenges associated with crop planning and yield estimation.

Yogesh Gandge et al [3] has explored the application of various data mining techniques for predicting crop yield, emphasizing that the success of such systems largely depends on the accuracy of feature extraction and the suitability of the classification algorithms employed. This thesis presents a comprehensive review of findings from multiple studies, highlighting the performance of various algorithms used by researchers in crop yield prediction. The analysis includes accuracy metrics and practical recommendations, providing

insights into the strengths and limitations of each method in real-world agricultural contexts.

Ramesh et al [4] focuses on applying data mining techniques to improve crop yield prediction accuracy. The authors analyze historical agricultural data and employ classification algorithms—specifically Decision Trees and Naïve Bayes—to identify patterns that influence crop production. The research emphasizes the role of environmental and soil parameters such as temperature, rainfall, humidity, and soil fertility in yield prediction. By comparing algorithmic performance, the paper demonstrates that data mining methods can support effective decision-making in agriculture, helping farmers choose the most suitable crops and optimize farming practices. The study concludes that integrating data mining into agricultural systems enhances prediction reliability and provides valuable insights into improving crop planning strategies.

Manjula et al [5] proposes a crop yield prediction model using data mining and machine learning techniques, with an emphasis on improving accuracy and reliability. The authors utilize regression analysis and classification methods to analyze historical agricultural data, including variables such as rainfall, temperature, soil type, and crop characteristics. The model is designed to assist farmers in selecting appropriate crops and planning agricultural activities based on predictive insights. The study highlights the effectiveness of combining multiple parameters and algorithms to achieve better prediction outcomes. It concludes that data-driven models can play a critical role in enhancing agricultural productivity and decision-making.

Raorane et al [7] explores the application of data mining techniques as a powerful tool for crop yield estimation in the agricultural sector. The authors emphasize the potential of data mining to uncover hidden patterns and correlations within historical agricultural datasets. Techniques such as classification, clustering, and association rule mining are discussed in the context of analyzing factors like crop type, soil condition, climate data, and historical yield records. The study advocates for the integration of these techniques to support strategic agricultural planning, improve resource allocation, and enhance productivity. The paper concludes that data mining serves as a vital decision-support tool for farmers and policymakers by enabling more informed, data-driven agricultural practices.

Table1. Data Set Description

Variables	Description
District	The data was collected 31 district of Tamil Nadu in India
Year	The data was taken from the year 2015 -2016
Crop	Crop like chili, onion, banana, potato, tomato, etc...
Area	The total area of agriculture plant region in Hectares.
Production	The production of crop in the specified year in metric tons.
Soil Type	Soil type like alluvial, red loamy, sandy loam etc...
Variables	Description
Soil fertility	Soil fertility in micro nutrients are Zn, Fe, Cu, MN,B and S
Rainfall	Rainfall data was collected 31 district of Tamil Nadu in India.

DATASET

To start with any data mining problem, it is first necessary to bring all data together. The dataset used in this research has been collected from data.gov.in. The data is used for this proposed work for district of Tamil Nadu in India. The data taken in eight input variables. The variables are year, district, crop, soil type, soil nutrient, area, rainfall and production.

DATA TRANSFORMATION

Using data transformation method in dataset to convert nominal to numerical data.

DATA DISCRETIZATION

In dataset the attribute production value is converted into very low, low, medium and high.

DATA MINING

Data mining is the process of looking at huge set of information to generate new information. Data mining is all about discovering unsuspected / previously unknown relationships amongst the data.

DECISION TREE FOR REGRESSION

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values. This model to predict only continuous value. The label attribute is already in numerical value so did it change, now to select the criterion measure is least square method. Least square is a statistical procedure to find the best fit for a set of data points. Least squares regression is used to predict the behavior of dependent variables.

DECISION TREE FOR CLASSIFICATION

Classification help to classify data in different classes. The label attribute which is continuous is discretized to very low, low, medium and high using discretize operator. Now the label attribute becomes categorical. Using decision tree for classification to select the criterion measure is Gain ratio method. Gain ratio method reduces the bias towards multivalued attribute by taking the number and size of branch into account when choosing an attribute.

$$Gain\ Ratio\ (A) = Gain(A) / Split\ info\ (A)$$

RANDOM FOREST FOR REGRESSION

Using Random forest regression in Rapid miner studio. In regression model to predict only numerical value. Using this operator to set the No.of tree=100 criterion measure used is least square method. Least square is a statistical procedure to find the best fit for a set of data points. Least squares regression is used to predict the behavior of dependent variables.

Table2. Performance of crop production

Performance parameters	Decision Tree	Random Forest
Absolute error	948.858	1740.437
Normalized absolute error	0.089	0.164
Squared correlation	0.954	0.958

ANALYSIS AND RESULTS

The proposed method use past year dataset to predict future crop production. Based on those algorithm the best performance model is decision tree regression.

Here selected the major factors like micro soil nutrients and rainfall, based on this factors to predict the crop production for future year with the help of decision tree regression model.

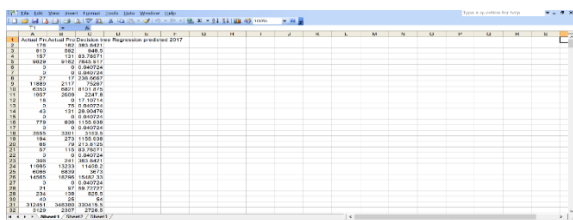


Fig1. Predicted values using decision tree regression

ALGORITHM

Using past year dataset, to predict crop production in various district in Tamil Nadu based on district which crop produce more production. Here mention some of district are:

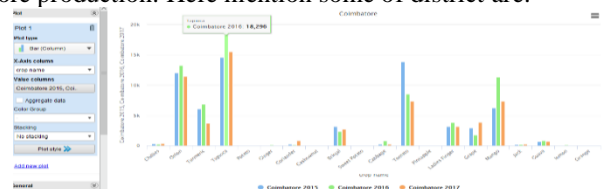


Fig2. Crop Production In Coimbatore District

In above Fig2. graph show that major crop produce in Coimbatore district,the crops are Tapioca, onion, tomato and mango.

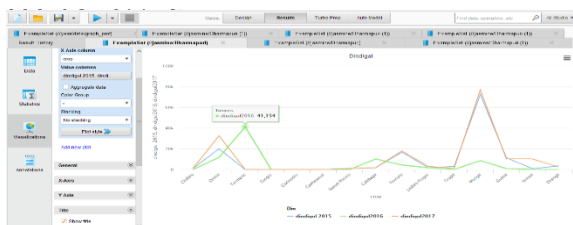


Fig3. Crop Production In Dindigul District

In above Fig3. graph show that major crop produce in Dindigul district,the crops are Guava, ginger and onion.

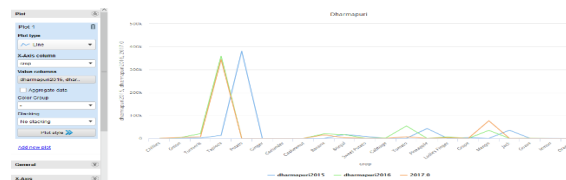


Fig4. Dharmapuri District Crop Production

In above Fig4. graph show that major crop produce in Dharmapuri district,the crops are tapioca, turmeric, potato and ginger.

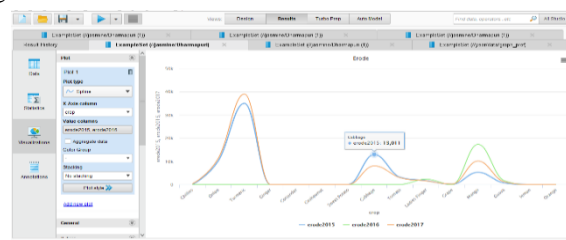


Fig (5) Erode District Crop Production

In above Fig(5) graph show that major crop produce in Erode district,the crops are chillies, cginger, sweet potata and guava.

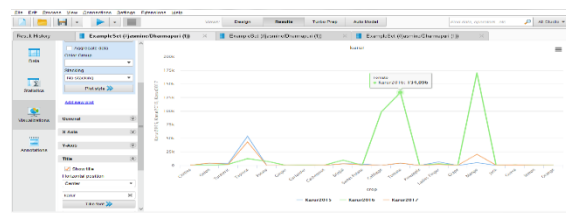


Fig (6) Crop Production In Karur District

In above Fig(6) graph show that major crop produce in Karur district,the crops are turmeric, potato and mango.

FUTURE WORK

In future it can be observed from above there are still many challenges in this research area using another types of crops like food crops, cash crops and plantation crops with macro soil nutrients to predict the crop yield.

CONCLUSION

In this work, production of horticultural crops and minor crops in 31 districts of Tamil Nadu were analyzed and classified as very low, low, medium and high using decision tree years and the micro soil nutrient of Zinc(Zn), Iron(Fe), Copper(Cu), Manganese(Mn), years and the micro soil nutrient of Zinc(Zn), Iron(Fe), Copper(Cu), Manganese(Mn), Boron(B) and Sulphur (S) as major contributing factors. The data for the year 2015 & 2016 are taken as training set for the model and crop yield for year 2017 has been predicted. Boron(B) and Sulphur (S) as major contributing factors. The data for the year 2015 & 2016 are taken as training set for the model and crop yield for year 2017 has been predicted algorithm. A prediction

model was developed by considering rainfall data of past Decision tree and Random forest algorithm are used for prediction, the two algorithm are compared. It was observed that Decision tree for regression algorithm produces better performance. The result are visualized for clarity of results. This work is a pioneering one in the agricultural scenario, for it uses data mining techniques to classify and predict possible crop yield district-wise in Tamil Nadu, based on soil content and rainfall. This work could be of significant use to farmers, for it uses technology to predict possible yields of various crops in specific regions of the state of Tamil Nadu, which could critically aid the farmer when he makes his decision regarding the kind of crop to sow in his field in the upcoming cycle. Further, this model can also be extended to other parts of India to predict crop yield.

REFERENCES

- [1] Jiawei, H., & Micheline, K. (2006). Data mining: concepts and techniques. Morgan kaufmann.
- [2] P.S. Vijayabaskar, Sreemathi.R, Keerthana E “Crop Prediction Using Predictive Analytics”, International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), 2017
- [3] Yogesh Gadge, Sandhya “A Study on Various Data Mining Techniques for Crop Yield Prediction”, International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECOT), 2017
- [4] D Ramesh , B Vishnu Vardhan “Analysis of Crop Yield Prediction Using Data Mining Techniques”, IJRET: International Journal of Research in Engineering and Technology, 2015
- [5] E. Manjula S. Djodiltachoumy “A Model for Prediction of Crop Yield”, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017
- [6] Ramesh A. Medar, Vijay. S. Rajpurohit “A survey on Data Mining Techniques for Crop Yield Prediction”, ISSN: 23217782 (Online) Volume 2, Issue 9, September 2014 International Journal of Advance Research in Computer Science and Management Studies
- [7] Raorane A.A.1 , Kulkarni R.V.2 “Data Mining: An effective tool for yield estimation in the agricultural sector”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) ,2012