International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 03 | March - 2021 ISSN: 2582-3930

CROP YIELD PREDICTION BASED ON INDIAN AGRICULTURE USING MACHINE LEARNING

DURGA DEVI S
Asst. Prof. Mr. S. BARATH
KRISHNASAMY COLLEGE OF ENGINEERING AND TECHNOLOGY

Abstract: Agricultural statistics and forecast is an important resource that the government has not explored commensurate to its impact. The aim of this project is to make this process computerized by implementing principles of data mining and analytics. More specifically, this project aims at targeting the social issue of drought, analyzing data based on crop produce, amount of rainfall, agricultural inputs, irrigation, and similar factors for every crop in the state of Tamil Nadu. Based on the extensive research carried out through this project, effective counter measures and suggestions will be given, which if implemented expeditiously, can help tackling the problem of drought in our state. Data can be mined and analyzed to find various trends and relations, such as – contrast between total irrigation area and type of crop; total principal and non-principal crop amount versus district-wise rainfall etc. The end result of the project will be research based reports specifying these trends, studied and analyzed from data taken over the past few years. Actions to minimize the damage of drought will also be suggested.

Keywords: Machine Learning, Crop yield prediction, Decision tree, SVM, Rainfall prediction, Forecasting, Data Mining.

1. Introduction

In the current scenario, the government is collecting data only in its raw form, and this data is of no use to the end user, that is the farmers. Collecting this raw data, standardizing it, analyzing it, and feeding it to a system that will provide relational trends is the aim of this project.

These relational trends will act as solutions for farmers, especially in drought afflicted areas. For example, the cultivation of Kharif Rice is resource intensive and should be only done in a rainfall rich period, otherwise it could use up the natural reserves of ground water, leading to deficiency in the water table, which consequently leads to drought like conditions.

This example, as naïve as it may sound, represents a broad class of trends which when extrapolated efficiently using the existing data mining algorithms, can produce a plethora of solutions, which when adhered to, will help alleviate the aforementioned drought like conditions.

The ultimate objective of this project is to stand as a system that when fed with data regarding various parameters, successfully produces trends and correlations that can help the user of the system develop solutions to tackle or minimize the damage of drought. Computerization of this process will drastically reduce the time required to study patterns and carry out extensive research to generate reports, and will give a close estimation of the required outcome.

2. Literature Review

a) Predicting yield of the crop using machine learning algorithm. This paper focuses on predicting the yield of the crop based on the existing data by using Random Forest algorithm. Real data of Tamil Nadu were used for building the models and the models were tested with samples. Random Forest Algorithm can be used for accurate crop yield prediction.

b) Random forests for global and regional crop yield prediction. This generated outputs show that RF is an effective and adaptable machine-learning method for crop yield predictions at regional and global scales for its high accuracy and precision, ease of use, and utility in data analysis. Random Forest is the most efficient strategy and it outperforms multiple linear regressions (MLR).

c) Machine learning approach for forecasting crop yield based on parameters of climate. The paper provided in International Conference on Computer Communication and Informatics (ICCCI). In the current research a software tool named Crop Advisor has been developed as a user friendly web page for predicting the influence of climatic parameters on the crop yields.C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. The paper is implemented using Decision Tree.

d) Prediction on Crop Cultivation. Presently, soil analysis and interpretation of soil test results is paper based. This in one way or another has contributed to poor interpretation of soil test results which has resulted into poor recommendation of crops, soil amendments and fertilizers to farmers thus leading to poor crop yields, micro-nutrient deficiencies in soil and excessive or less application of fertilizers. Formulae to Match

Crops with Soil, Fertilizer Recommendation.

e) Analysis of Crop Yield Prediction by making Use Data Mining Methods. In this paper the main aim is to create a user- friendly interface for farmers, which gives the analysis of rice production based on the available data. For maximizing the crop productivity various Data mining techniques were used to predict the crop yield. Such as K-Means algorithm to forecast the pollution factor in the atmosphere.

f) Applications of Machine Learning Techniques in Agricultural Crop Production. From GPS based color images is provided as an intensified indistinct cluster analysis for classifying plants, soil and residue regions of interest. The paper includes various parameters which can help the crop yield for better enhancement and ratio of the yield can be increased during cultivation.

g) In this paper, we present a comprehensive review of research dedicated to the application of machine learning in agricultural production systems. Machine learning (ML) has emerged together with big data technologies, techniques, methods and high-performance computing to generate new opportunities to unravel, quantify, and analyze data intensive processes in agricultural operational sectors. By using Support Vector Machines (SVP) the Paper is implemented.

h) Random Forests for Global and Regional Crop Yield Predictions. The generated outputs show that RF is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy. The Paper is Implemented using k-nearest neighbor, Support Vector Regression (SVG).

3. Proposed System

In the proposed system, we use supervised learning to form a model, which provides predicted cost of crop yield and corresponding production order. The proposed system is described in following stages such as dataset collection, preprocessing step, feature selection and applying machine learning modules as shown in figure1.

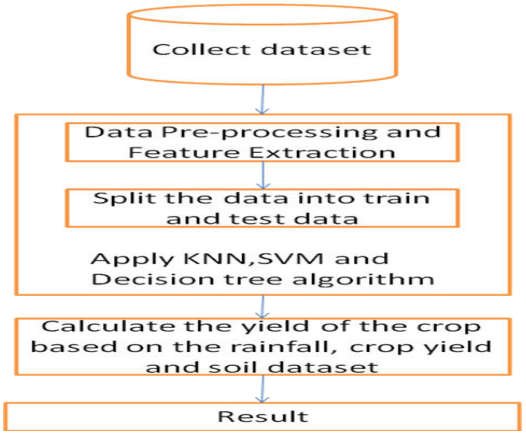

Figure.1 System Architecture

A. Dataset Collection: Data is collected from a variety of sources and prepared for data sets. And this data is used for descriptive analysis. Data is available from several online abstract sources such as Kaggle.comand data.gov.in.

B. Preprocessing step: This step is a very important step in machine learning. Preprocessing consists of inserting themissing values, the appropriate data range, and extracting the functionality. The kind of the dataset is critical to the analysis process. In this paper we have used isnull() method for checking null values and label Encoder() for converting the categorical data into numerical data.

C. Feature Selection: Feature extraction should simplify the amount of data involved to represent a large data set. The soiland crop characteristics extracted from the pre-treatment phase constitute the final set of training. These characteristics include the physical and chemical properties of the soil. Here, we have used Random Forest Classifier () method for feature selection. This method selects the features based on the entropy value i.e., the attribute which is having more entropy value is selected as important feature for yield prediction.

D. Split the Dataset into Train and Test Set: This step includes training and testing of input data. The loaded data is divided into two sets, such as training data and test data, with a division ratio of 80% or 20%, such as 0.8 or 0.2. In a learning set, a classifier is used to form the available input data. In this step, create the classifier's support data and preconceptions to approximate and classify the function. During the test phase, the data is tested. The final data is formed during preprocessing and is processed by the machine learning module.

E. Apply Machine Learning Techniques: In this project, different supervised machine learning techniques for prediction of crop yield are used which is given as follows in Figure 2

Framework for Crop Yield Prediction

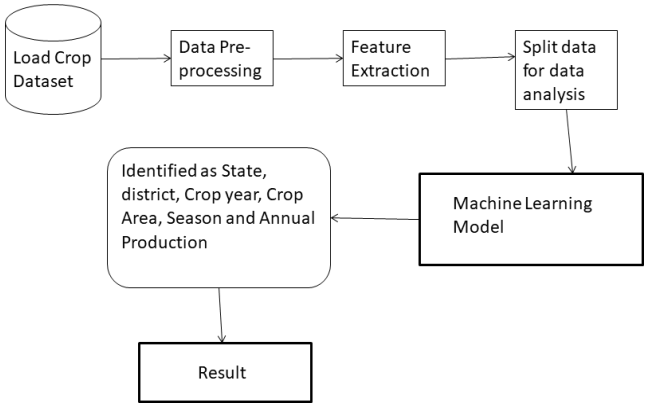

Figure.2 Framework for Crop Yield Prediction

4. Conclusion

Presently farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of crop for cultivation that will reduce their income. To reduce those type of loses, a farmer friendly system with GUI, that will predict which would be the best suitable crop for particular land and this system will also provide information about required nutrients to add up, required seeds for cultivation, expected yield and market price. So, this makes the farmers to take right decision in selecting the crop for cultivation such that agricultural sector will be developed by innovative idea

5. Future Scope

To collect all required data by giving GPS locations of a land and by taking access from Rain forecasting system ofby the government, we can predict crops by just giving GPS location. Also, develop the model to avoid over and under crisis of the food.

6. References

[1] V. Ramesh and K. Ramar, "Classification of Agricultural Land Soils: A Data Mining Approach", *Agricultural Journal*, vol. 6, no. 3, pp. 82-86, 2011.

[2] P. Bhargavi and S. Jyothi, "Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils", *International Journal of Computer Science and Network Security*, vol. 9, no. 8, pp. 117-122, 2009.

[3] G. Yi-Yang and R. Nan-ping, "Data Mining And Analysis Of Our Agriculture Based On The Decision Tree", *International Colloquium on Computing, Communication, Control, and Management*, pp. 134-138, 2009.

[4] D. Rajesh, "Application of Spatial Data mining for Agriculture", *International Journal of Computer Applications*, vol. 15, no. 2, pp. 7-9, 2011.

[5] N. Hemageetha, "A Survey on Application of Data Mining Techniques to Analyze the Soil for Agricultural Purpose", *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi*, pp. 3112-3117, 2016.

[6] R. Sujatha and P. Isakki, "A Study of Crop Yield Forecasting Using Classification Techniques", *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti*, pp. 1-4, 2016.

[7] D. Ramesh, "ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING

TECHNIQUES", *International Journal of Research in Engineering and Technology*, vol. 04, no. 01, pp. 470-473, 2015.

[8] G. Fathima and R. Geetha, "Agriculture Crop Pattern Using Data Mining

Techniques", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 5, pp. 781-786, 2014.

[9] H. Patel and D. Patel, "A Brief survey of Data Mining Techniques Applied to Agricultural Data", *International Journal of Computer Applications*, vol. 95, no. 9, pp. 6-8, 2014.

[10] "Introduction to Support Vector Machines — OpenCV 2.4.13.2 documentation", *Docs.opencv.org*. [Online]. Available:

http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html. [Accessed: 17- Apr- 2019].