International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228681

Design A Model for Crop Prediction and Analysis Using Machine Learning

Ruchira C. Mahore[1], Naresh G. Gadge[2]
[1]MTech, Department of Computer Science Govt. College of Engineering, Amravati, India
[2] Assistant Professor, Department of Computer Science Govt. College of Engineering, Amravati, India

Article Info
Publication Issue :
Volume 9, Issue 1
January-February-2023
Page Number : 90-93

Article History
Accepted: 01 Jan 2023
Published: 20 Jan 2023

ABSTRACT

Agriculture is the most important aspect of survival. Machine learning (ML) could be an important point of view in determining a practical and workable solution to the crop yield problem. Given the current method, which includes manual counting, climate-smart pest management, and satellite photography, the results aren't particularly accurate. The primary goal of this research is to forecast crop and yield yield using various machine learning approaches. SVM, Nave Bayes, and Random Forest are the classifier models used here, with Random Forest providing the highest accuracy. Machine learning algorithms will help farmers choose which crop to cultivate based on variables such as temperature, rainfall, area, and other characteristics This connects the technological and agricultural sectors .

Keywords : Crop Prediction, SVM, Machine Learning

I. INTRODUCTION

AI calculations are utilized in information mining to analyze required information in the suitable manner. Since quite a while in the past, AI has been utilized in the rural area [8]. A great deal of AI procedures is being utilized to gauge horticultural yield methodologies. Since crop creation forecast is perhaps of the most difficult test in the agrarian area, a few models have been proposed to date [1], [16]. This undertaking utilizes a wide assortment of datasets on the grounds that the yield of a harvest relies upon various variables, including the country area, the yield's name, the year, the yield esteem, how much normal precipitation, the utilization of pesticides, and the typical temperature [1]. A subset of computerized reasoning is AI. A reasonable and deliberate system can assist us expect crop yields all the more precisely as per different meteorological factors. The datasets are used to prepare the model that begins the result in view of authentic information [4]. The dataset should be isolated into preparing and testing segments for expectation purposes [3]. A preparation dataset is utilized to construct a model in a dataset, and a testing dataset is utilized to approve the model that was created utilizing the preparation dataset [9]. This shows that foreseeing horticultural creation is a vital and significant errand given the wide assortment of harvests filled in various countries [2].

II. LITERATURE SURVEY

According to Dr. Y. Jeevan Nagendra Kumar et al. [1], numerous data mining approaches can be used to increase accuracy rate. The dataset includes variables such as temperature, rainfall, humidity, and pH. A massive number of decision trees are created in a training scenario, and the outcome or output is divided using the number of classes. To compare the two and choose the better one, this paper also utilizes a decision tree classifier. They forecasted the outcome using supervised learning techniques. To train the model, they contrasted Random Forest and Decision Tree. Limitation: This paper does not employ assembling methods. Judah Immaculate and others The machine learning algorithms used in the agriculture sector are described in [2]. This calculates the efficacy of each machine learning system that can forecast agricultural yield. This also discusses how the algorithm is implemented and how it functions. The efficiency of every algorithm utilised is calculated in this publication. This also explains about how effectively the machine learning approaches are suitable for prediction of crop yield. Limitation: Not all algorithms are described in full in this study, nor are assembling strategies employed. Anuja Changude and others [3], introduces a system that uses sensors to retrieve data. Additionally, data prediction is accomplished using the ANN (Artificial Neural Networks) machine learning technique. The dataset includes variables like temperature, wind speed, humidity, and sunlight. Here, machine learning is used to forecast crop growth and diseases. More emphasis is placed on monitoring using IOT devices in this paper. Limitation: Rather than predicting, this article focuses primarily on monitoring with IOT devices. Pavan Patil et al. [4] offers an efficient crop recommendation system utilizing classifier. Decision Tree and KNN, two ML classification algorithms, were compared in this work. The data set consists of productivity, meteorological, and soil factors. Algorithms for machine learning are used to compute in this research. This paper compared the two algorithms separately without combining them. This is an example of a recommendation system that uses the author-appropriate classifiers. Limitation: Ensembling techniques are not used, therefore KNN provides lower accuracy than other algorithms. Saeed Khaki and others [5], analysed the yield, compared the yield difference, and forecasted yield using neural networks. Additionally, they calculated the algorithm's efficiency and discussed how well it is suited. This document provides more accurate information and effective outcomes. However, neural networks make the implementation more complex, and theses function better with image data than with numerical data. Limitation: Neural networks are more difficult to implement and are better suited for processing picture data than numerical data. K.Ruth Ramya and others [6] explains how to predict agricultural productivity using a random forest classifier. They used minimum temperature, maximum temperature, humidity, and rainfall as parameters; the results were 99.7% accurate. The only drawback of this paper is the lack of an ensemble algorithm, which lowers the accuracy in comparison. Rohan Yadav and others [7], The Deep Learning Approach is the major topic of this study. They have taken factors like pH, nitrogen, phosphate, copper, etc. into consideration. More hidden layers could improve the results' accuracy. Based on a few measures, they compared the accuracy of all neural network methods. To eventually forecast the result, they employed DNN. Limitation: The implementation of neural networks is more complicated. Priya and others [8] focus mainly on supervised learning techniques like decision tree and random forest classifiers. The datasets in this case include more than half of the records in the datasets and include rainfall, perception, production, temperature, and a variety of decision trees. Following classification, decision trees are applied to the remaining records in order to boost accuracy. We learned from this paper that using RF to assemble a decision tree may produce more accurate and better

results. Limitation: Ensembling may employ more complicated algorithms.

III. METHODOLOGY

In this paper, crop datasets from government websites are collected, general data analysis is performed, followed by exploratory analysis (year, season, area, production, etc.). Correlation matrices are generated, an svm model is trained and tested, and results are produced using the Naive Bayes algorithm. Python programming is used throughout. In this paper, NumPy and library pandas are used. Scikit-learn, Matplotlib, Seaborn. A system called crop yield prediction is used to forecast the ideal crop based on the variables taken into account in the dataset. A combination of algorithms quickly increases accuracy. The following modules make up the process: Pre-processing: There are two categorical columns in the dataset. Categorical data are attributes with values that are presented in a labelled fashion. In this scenario, the values for the crop name and nation fall under a fixed set that doesn't go beyond the range of possible values. Direct use of labelled data by several machine learning algorithms is not possible. Data must be transformed into numerical data using one of the encoding techniques in order to operate the label. Data exploration involves looking at the connections between the dataset's columns. Using a heat map to visualise the correlation matrix is the best technique to examine correlations between columns. Pearson The correlation matrix is displayed as a heat map using correlation. Scaling: The range, magnitudes, and units of the properties in the data frame are all significantly different. As opposed to characteristics with smaller values, those with high values will provide a great deal of complexity and confusion. We must bring all features to the same range of magnitudes in order to overcome this. To achieve this, scale the data using the Standard Scaler or Min Max Scaler. Phase of Training and Testing: The training and testing data are a key component of the data preparation process. The most frequently thought-of ratio was taken into consideration in this case: 70/30. This forecasts how successfully the test data will be taught to produce the most accurate result. Comparison and selection of models: Before picking the method to apply, evaluation should be conducted to see which one is best suited for the particular dataset. Basically, we test different models to address optimization problems when working on a machine learning problem with a given dataset. However, the appropriate model won't over- or under-fit the data. A prediction method that establishes the relationship between a dependent and independent variable is regression analysis. Here, we'll employ two distinct techniques. The Naive Bayes method comes in first, followed by the Support Vector Machine approach.

Naive Bayes Algorithm

The foundation of the Naive Bayes algorithm is the Naive Bayes classifier. This classifier assists in calculating the likelihood of the anticipated classes. Large datasets can be easily built using this technique.

Support Vector Machine (SVM)

The foundation of the Naive Bayes algorithm is the Naive Bayes classifier. This classifier assists in calculating the likelihood of the anticipated classes. Large datasets can be easily built using this technique .Any number of hyperplanes may be drawn, but the algorithm's primary goal is to locate the plane with the largest margin, or the greatest separation between the data points of the features being plotted. The classification will be more accurate the more away it is. The data points are separated from all other points in the optimal hyperplane. Creating a maximum margin. We chose Linear SVM in this proposed model because it was suitable for our kind of prediction.

Here is the dataset for the crop recommendation:

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.879774	82.00274	6.502985	202.9355	rice
71	54	16	22.6136	63.69071	5.749914	87.75954	maize
40	72	77	17.02498	16.98861	7.485996	88.55123	chickpea
13	60	25	17.13693	20.59542	5.685972	128.2569	Kidney beans
3	72	24	36.51268	57.92887	6.031608	122.654	Pigeon peas
3	49	18	27.91095	64.70931	3.692864	32.67892	mothbeans
19	55	20	27.43329	87.80508	7.185301	54.73368	mungbean
56	79	15	29.4844	63.19915	7.454532	71.89091	blackgram

IV. CONCLUSION AND FUTURE SCOPE

Finally, based on previously gathered data, a system for crop yield prediction has been put into place. With the help of some machine learning algorithms, this has been resolved. To forecast the outcome with a higher rate of accuracy, an ensemble of decision tree regression and an AdaBoost regression is utilized in this case. As opposed to other algorithms, decision trees will improve accuracy, as we have realized. Decision trees by themselves do not produce many correct results, which leads to a subpar output. AdaBoost regressor is ensembled to strengthen the decision tree's weak learner. Increased accuracy will be more beneficial to the yield. Based on the parameters, this technique helps farmers choose the crop that should be grown in the land.

V. REFERENCES

[1]. Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 473-478). IEEE.

[2]. Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846). Springer, Berlin, Heidelberg.

[3]. Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019, August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE.

[4]. Bhosale, S. V., Thombare, R. A., Dhemey, P. G., & Chaudhari, A. N. (2018, August). Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[5]. Gandge, Y. (2017, December). A study on various data mining techniques for crop yield prediction. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) (pp. 420-423). IEEE.

[6]. Gandhi, N., Petkar, O., & Armstrong, L. J. (2016, July). Rice crop yield prediction using artificial neural networks. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 105-110). IEEE.

[7]. Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016, July). Rice crop yield prediction in India using support vector machines. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 1-5). IEEE.

[8]. Gandhi, N., Armstrong, L. J., & Petkar, O. (2016, July). Proposed decision support system (DSS) for Indian rice crop yield prediction. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 13-18). IEEE.

[9]. Islam, T., Chisty, T. A., & Chakrabarty, A. (2018, December). A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1-6). IEEE.

[10]. Jaikla, R., Auephanwiriyakul, S., & Jintrawet, A. (2008, May). Rice yield prediction using a support vector regression method. In 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (Vol. 1, pp. 29-32). IEEE.

[11]. Kadir, M. K. A., Ayob, M. Z., & Miniappan, N. (2014, August). Wheat yield prediction: Artificial neural network-based approach. In 2014 4th International Conference on Engineering Technology and Technopreneurs (ICE2T) (pp. 161-165). IEEE.

[12]. Manjula, A., & Narsimha, G. (2015, January). XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. In 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO) (pp. 1-5). IEEE.

[13]. Mariappan, A. K., & Das, J. A. B. (2017, April). A paradigm for rice yield prediction in Tamilnadu. In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 18-21). IEEE.

[14]. Paul, M., Vishwakarma, S. K., & Verma, A. (2015, December). Analysis of soil behaviour and prediction of crop yield using data mining approach. In 2015 International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 766-771). IEEE.

[15]. Shah, A., Dubey, A., Hemnani, V., Gala, D., & Kalbande, D. R. (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. In Proceedings of International Conference on Wireless Communication (pp. 49-56). Springer, Singapore.

[16]. Ahamed, A. M. S., Mahmood, N. T., Hossain, N., Kabir, M. T., Das, K., Rahman, F., & Rahman, R. M. (2015, June). Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 1-6). IEEE.

[17]. Shastry, A., Sanjay, H. A., & Hegde, M. (2015, June). A parameter based ANFIS model for crop yield prediction. In 2015 IEEE International Advance Computing Conference (IACC) (pp. 253-257). IEEE.

[18]. Sujatha, R., & Isakki, P. (2016, January). A study on crop yield forecasting using classification techniques. In 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp. 1-4). IEEE.

[19]. Suresh, A., Kumar, P. G., & Ramalatha, M. (2018, October). Prediction of major crop yields of Tamilnadu using K-means and Modified KNN. In 2018 3rd International Conference on Communication and Electronics Syst ems (ICCES) (pp. 88-93). IEEE.

[20]. Veenadhari, S., Misra, B., & Singh, C. D. (2014, January). Machine learning approach for forecasting crop yield based on climatic

parameters. In 2014 International Conference on Computer Communication and Informatics (pp. 1-5). IEEE.

Cite this article as :

Ruchira C. Mahore, Naresh G. Gadge, "Design A Model for Crop Prediction And Analysis Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9 Issue 1, pp. 90-95, January-February 2023. Available at doi :
https://doi.org/10.32628/CSEIT228681
Journal URL : https://ijsrcseit.com/CSEIT228681