IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

Crop Yield Prediction Using Machine Learning

Dr Madhusudhan S[1], Kiran A S[2], Vikas M[3], Praveen Kumar[4], Sandesh Kumar KM[5]
Associate Professor, Department of Computer Science and Engineering[1]
UG Students, Department of Computer Science and Engineering[2,3,4,5]
Alva's Institute of Engineering and Technology, Tenkamijar, Karnataka, India

Abstract: *India is an agricultural country and has such vast arable land, but yields are still very poor. There are many reasons for this. Basically, agriculture is dependent on monsoons, etc. However, one of the serious problems is that farmers are not aware of the latest technology and its use. Our project is mainly focused on predicting suitable crops for a specific time period area. This forecasting process depends primarily on meteorological conditions (monsoon) soil type, soil nutrientlevel, soil moisture level, etc. In our project, the model takes the above parameters as input and predicts suitable crops to grow in that environment. Depending on the values, the dataset can be prepared and trained using a machine-learning algorithm to achieve a specific output.*

Keywords: Soil Moisture, Temperature, Datasets,Humidity, Monsoon.

I. INTRODUCTION

Agriculture is the backbone of any economy. In countries such as India, where the demand for food is increasing due to population growth, the development of the agricultural sector is required to meet that demand. In ancient times, agriculture was considered the main and most importantculture practiced in India. Ancient peoples have responded to their needs by growing crops on their land. Therefore, it is cultivated as a natural crop and has been used by many living things such as humans, animals, and birds. Green products produced in the land brought by living things lead to a healthy and welfare life. Since the invention of new and innovative technologies and technologies, the agricultural sector has beenslowly deteriorating. As a result, many inventions are concentrated on the cultivation of hybrid product artificial products that bring about a life harmful to health. Today,modern people are unaware of growing crops at the right time and place. These cultivation techniques change seasonal and climatic conditions for basic assets such as soil, water and air, leading to food instability. By analyzing all these problemsand the problems of weather, temperature, many factors, etc. there is no proper solution or technique to overcome the situation we are facing. In India, there are several ways to stimulate economic growth in the agricultural sector. There aremany ways to improve and improve the yield and quality of crops.

Implementation of machine learning technology could have a big profit impact in agriculture. Predictions made by machine learning algorithms help farmers decide which crops to grow for maximum yield. The growth of agriculture depends on meteorological aspects such as temperature, rainfall, humidity,etc. rather than the surface temperature of the rotation of various soil parameters. The creation of this dataset has been reviewed in India over the past decade to ensure the learning and training of algorithms, Increases the accuracy rate of predictions. Yield forecasting is an important agricultural problem. Each farmer always wants to know how much harvest he can get from his expectations. In the past, yield forecasts were calculated by analyzing the farmer's previous experience with a particular crop. Agricultural yields depend mainly on weather conditions, pest harvesting work plans.Accurate information about the history of yields is critical for decisions related to agricultural risk management. Farmers check the yield of their crops per acre before planting in their fields.

II. RELATED WORK

J.P. Singh, Rakesh Kumar, M.P. Singh and Prabhat Kumar have come to the conclusion that a paper on applying classification methods to compare parameters and improve theprobability of crop production will be helpful. You can also use the Bayesian algorithm to analyze and predict crops. The algorithms used are Bayesian algorithms, Kmeans algorithms are clustering algorithms, and support vector machines. The disadvantage is the lack of proper accuracy and performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

According to Subhadra Mishra authors, Debahuti Mishra and Gour Hari Santra, came that it is an advanced research field and will grow in the future. The integration of computer science with agriculture helps to share agricultural harvest.

This method also helps to provide harvest information and increase productivity rates. The algorithms used are artificial neural networks, decision tree algorithms, regression analysis. The disadvantage is that the clear method is not displayed.

Author Karan Deep Kauri, concluded that his article part that various applications of machines are taken into account in the agricultural sector. And also provides insight into the problems in which Indian farmers face and how they can be solved with these techniques. This method helps to increasethe agricultural sector in countries and apply more learning applications. The algorithms used are artificial neural network,Bayesian networks, decision tree algorithms, clusters, regression analysis. The disadvantage is less accuracy of performance.

In E. Manjula Articles, S. Djodiltachoumy, concluded that the purpose of its article proposes and implement a rule-based system. And predictions produce crop yields from the previousdata collection. The algorithms used are Kmean's algorithms, clustering methods. Disadvantages are only suitable for theuse of linear rules and in consideration to consider less data.

According to the authors B.Mallicajun Rao, D.Sindhura, B.Navya Krishna, K.Sai Prasanna Lakshmi, Dr. J Rajendra Prasad came to the conclusion that this method provides a useful and accurate knowledge. With this knowledge, we forecast and support the decision-making for different fields. The algorithms used are many linear regressions. The disadvantage is that it can be applied to restricted areas.

In the article by T.Giri Babu, Dr.G.Anjan Babu, came the conclusion that the Agricultural algorithm method will supply solutions for farmers. You can also help to provide solutions for water and fertilizer problems. And this helps to get productive products. The disadvantage is that this method doesnot provide adequate accuracy for plants.

Authors Raorane A.A, Dr. Kulkarni R.V, concluded that this method will help to estimate rains and to investigate the reason to obtain lower yields. The algorithm used is theregression analysis method. The disadvantage is that the specific procedure is not displayed.

Siti Khirunnizabjo, Samihah Mustaffha, Wan Ishak Wan Ismail, concluded that this method offers a few peasant problems when providing good productivity. The algorithms used are artificial neuronal networks. The disadvantage is that it consumes more time.

According to international Technology Science and Technology Research technology, the research technique journal has on the prediction of system yields based on existing data with random forest algorithms. The Tamil Nadu data has been used to build models and models tested with samples. Random forest algorithms can be used to predictaccurate crop yields.

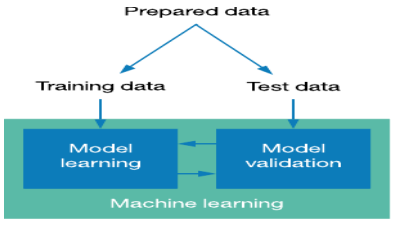
Random forests for the prediction of global and regional crop yields, according to the Institute for Environment, Minnesota University, St. Paul, MN 55108, USA. The output outputsshow that RF is an effective and different machine method to predict crop yields at regional and global scale on its high accuracy.

III. PROPOSED SYSTEM

CPython is the reference implementation of Python. Written in C, it meets the C89 standard with a selection of some C99 features. Compile your Python program with intermediate bytecode [101] and run it in a virtual machine. CPython is distributed with a large standard library created by mixing C and basic Python. It can be used on many platforms, including Windows and most modern Unixlike systems. Platform portability was one of the early priorities.

IV. METHODOLOGY

4.1 Modules



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

1. Data Pre-Processing
2. Train and Test Model

A. Data Pre-Processing

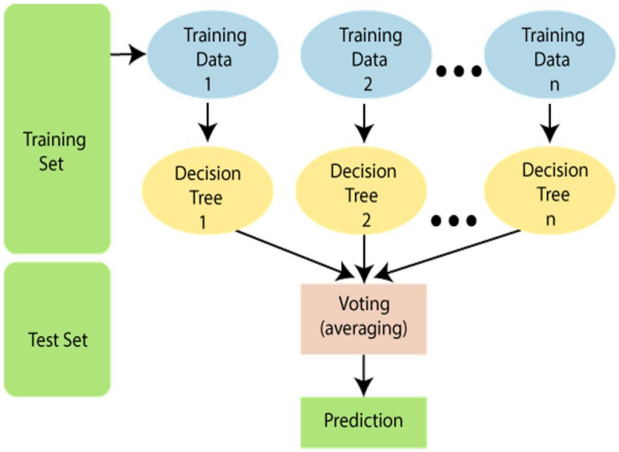
Grain yield forecast data is preprocessed after collecting various records. The dataset contains many records. The process of preparing raw data and adapting it to a machinelearning model. This is the first and important step in generating a machine learning model. The actual data generally contains noise, missing values, and has unusableformats that cannot be used directly in machine learning models. Data preprocessing is the work required to organize your data and adapt it to a machine learning model that improves the accuracy and efficiency of your machinelearning model.

Algorithm:

1. Converting the data sets into .csv format
2. Checking for null valuesData frame.isnull()

Syntax:

Pandas.isnull("DataFrame Name")or
DataFrame.isnull()
Parameters:Object to check null values.
Return Type: Data frame of Boolean values which areTrue or false values.



V. TRAIN AND TEST MODEL

Machine Learning can create a model that predictsthe outcome of a particular event and use a method called Train / Test to ensure that the model measures well. Training / testing is called because the method of measuring the accuracy of a model divides the dataset into two sets: a training set anda testing set 80% for training, and 20% for testing. Train your model using the training set. Test your model using the test set. Learning a model means creating a model. Testing the model means testing the accuracy of the model.

5.1 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problemsin ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier thatcontains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

The below diagram explains the working of the Random Forest algorithm:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

Random Forest works in two-phase first is to create therandom forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

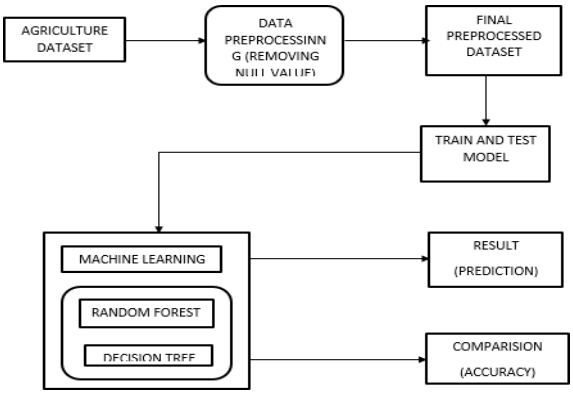
Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

5.2 Architecture

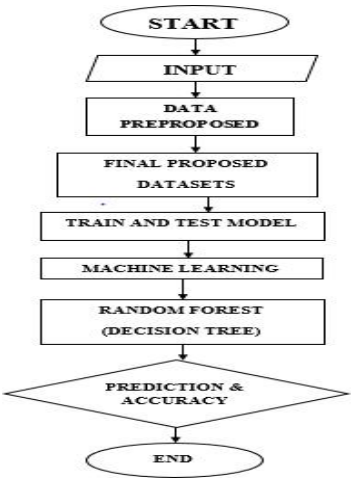


The figure above shows the design of a system generated with agricultural datasets as input, and these datasets are preprocessed with all null values removed. These preprocessed data sets are completed and passed to the training and test model. After learning and testing is complete,it will be sent to Machine Learning using the Random Forest Algorithm. For better accuracy, a decision tree is generated compared to the previous accuracy values. This prediction allows you to use machine learning to get the final result for a particular crop in a particular region or season.

5.3 Input /Output Design

I'm using a Python library flask to create a single GUI window. This GUI window allows users to provide conditions, crops, etc. in their area. Then, using the Random Forest Algorithm, the system predicts crop yields in a particular area and provides the accuracy of the trained model.

5.4 Flow Chart Design



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

5.5 Data Set

Datasets containing values of SO2, NO2, RSPM, SPM, AQI, RICE_AREA, RICE_PRODUCTION, ICE_YEILD, WHEAT_AREA,WHEAT_PRODUCTION,
WHEAT_YEILD choice tree algorithmic program. These datasets contain values of various eventualities within the fields so as to coach the model accurately. Sample datasets area unit as show in Table 1.

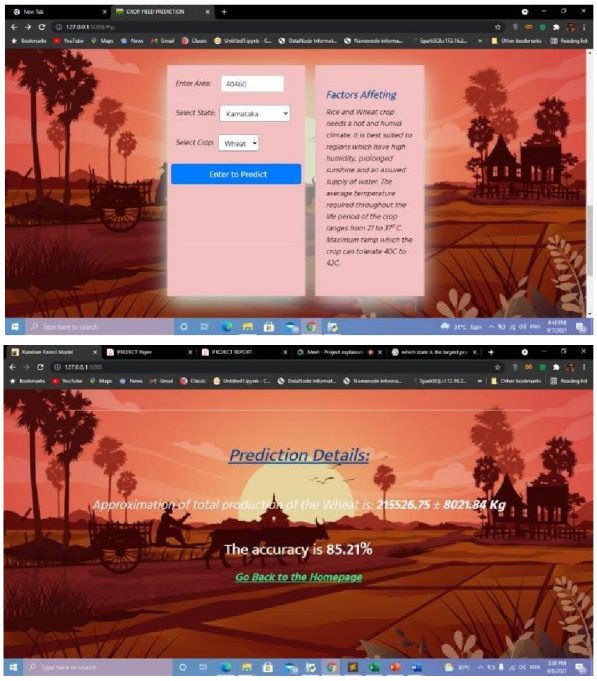
YEAR	STATE	SO2	NO2	RSPM	SPM	AQI	RICE_AREA	RICE_PRODUCTION	RICE_YIELD	WHEAT_AREA	WHEAT_PRODUCTION	WHEAT_YIELD
1998-99	KARNATAKA	7.45	18.55	140.261	318	318	56800	79220	1.39471831	85000	99600	1.171764706
1999-2000	KARNATAKA	10.01	15.46	140.261	322	322	60880	86000	1.41261498	90000	100000	1.111111111
2000-01	KARNATAKA	10.5	22.5	140.261	470	470	68615	79000	1.151351745	75480	90600	1.200317965
2001-02	KARNATAKA	10.2	38.46	140.261	335.85	335.85	63000	54500	0.865079365	48800	78550	1.609631148
2002-03	KARNATAKA	17.46	27.89	120.67	261.9	261.9	48920	54600	1.116107931	81500	103520	1.270184049
2003-04	KARNATAKA	18.45	41	120.67	452	452	43900	57000	1.298405467	75300	95200	1.264276228
2004-05	KARNATAKA	17.24	55.97	120.67	338	338	46098	54200	1.175755998	91650	139000	1.516639389
2005-06	KARNATAKA	16.5	55.51	135.64	425	425	45400	55900	1.231277533	64600	85500	1.323529412
2006-07	KARNATAKA	18.46	20.45	135.64	550	550	40580	53154	1.309857072	95600	165000	1.725941423
2007-08	KARNATAKA	16.45	25.98	135.64	460.23	460.23	60900	87120	1.430541872	95500	155500	1.628272251
2008-09	KARNATAKA	20.4	31.24	135.64	560	560	68615	75400	1.098885083	85000	112500	1.323529412
2009-10	KARNATAKA	22.86	42.06	135.64	356.25	356.25	46098	46000	0.997874094	84300	114800	1.361803084
2010-11	KARNATAKA	20.79	30.54	135.64	421.05	421.05	46000	55210	1.200217391	75000	96500	1.286666667
2011-12	KARNATAKA	22.36	47.56	135.64	575.29	575.29	60900	76100	1.249589491	84300	99000	1.174377224
2012-13	KARNATAKA	18.54	25.87	135.64	452	452	48500	55000	1.134020619	45700	78600	1.719912473
2013-14	KARNATAKA	23.45	28.45	135.64	255.9	255.9	44820	56210	1.254127622	75480	90510	1.199125596
2014-15	KARNATAKA	34.2	43.15	135.64	450	450	63000	87740	1.392698413	92000	139000	1.510869565
2015-16	KARNATAKA	19.24	35.79	135.64	512	512	45400	35561	0.783281938	81100	103510	1.276325524
2016-17	KARNATAKA	25.78	22.54	135.64	330	330	48500	51785	1.067731959	64600	85500	1.323529412
2017-18	KARNATAKA	38.22	58.02	135.64	542.32	542.32	60900	65500	1.075533662	56600	80900	1.429328622
2018-19	KARNATAKA	34.05	39.05	135.64	618	618	45400	45480	1.001762115	64580	86502	1.33945494
2019-20	KARNATAKA	33.25	21.03	135.64	360	360	47500	58500	1.231578947	78500	90000	1.146496815
2020-21	KARNATAKA	34.51	30.1	135.64	564.2	564.2	43950	43000	0.978384528	64580	84200	1.303809229

Fig 3: Datasets

VI. RESULT

The sample output as shown in figure four contains the values of temperature in each Centigrade and Gabriel Daniel Fahrenheit, humidity, water presence, associated prints likewise as sends an email aware of farmer.

.



1. Random Forest is flexible, easy to use Machine Learning algorithm.
2. Random Forest is a supervised learning algorithm.
3. The advantage of Random Forest Algorithm is Over fitting is less of an issue with Random Forests, unlike decision tree Machine Learning algorithms.
4. Random Forest Machine Learning Algorithms can be grown in parallel.
5. By applying the above algorithm we are getting the output for 10 acre is 215526 kgs.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

VII. CONCLUSION

We have proposed various machine learning algorithms for predicting crop yields based on temperature, rainfall, season and area. Experiments with the Government of India dataset have demonstrated that Random Forest Regressor offers the highest yield prediction accuracy. Other parameters such as rainfall, temperature, season, and area can be combined to predict output in a particular region. The results show that Random Forest is the best classifier when all the parameters are combined. This will not only help farmers choose the right crops to grow next season, but will also help close the gap between technology and the agricultural sector. The results show that accurate yield predictions can be achieved using the random forest algorithm. The random forest algorithm provides the model with the highest yield with the fewest models. Suitable for large yield forecasting in agricultural planning.

So farmers can make the right decisions about the right crops to advance the agricultural sector through innovative ideas. The value predicted by the algorithm is close to the value specified in the historical data. In developing countries like India, farmers play an important role in GDP. So, by developing ML architecture for agriculture, it enhances and contributes to its overall GDP. The system can be enhanced further to add following functionality:

1. By using IoT for taking inputs like soil moisture level and soil pH level sensors and temperature sensor by using the above sensor we can improve the accuracy
2. This paper describes crop yield prediction ability of the algorithm. In future we can determine the efficient algorithm based on their accuracy metrics that will helps to choose an efficient algorithm for crop yield prediction.
3. Crop diseases detection using Image Processing where users can upload picture of diseased crop and get pesticides recommendations.
4. Implementation of Smart Irrigation System to monitor weather and soil conditions, plant water usage etc. to automatically alter watering schedule.

REFERENCES

[1]. Mrs. K.R. Sri Preethaa M. E, S. Nishanthini, D. Santhiya, K. Vani Shree," CROP YIELD PREDICTION", Arasur, Coimbatore, International Journal On Engineering Technology and Sciences– IJETS, March- 2016

[2]. Vaneesbeer Singh, Abid Sarwar, "Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach" IJARCSE, vol. 5, Issue 8, 2017

[3]. http://agricoop.nic.in/sites/default/files/Annual_rpt_2 016 17_E.pdf

[4]. Aditya Shastry, H.A Sanjayand E.Bhanushree,"Prediction of crop yield using Regression Technique", International Journal of computing12 (2):96- 102 2017,ISSN:1816-9503

[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]. Technology and Sciences – IJETS™ISSN(P): 2349- 3968, ISSN (O):2349-3976 Volume III,Issue III, March-2016

[7]. Askar Choudhury, James Jones, "CROP YIELD PREDICTION USING TIME SERIESMODELS"

[8]. Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of agriculture data using datamining techniques: application of big data"Majumdar et al. J Big Data (2017) 4:20 DOI 10.1186/s40537-017-0077-4

[9]. D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015.

[10]. Yethiraj N G , " Applying data mining techniques in the field of Agriculture and allied sciences", Vol 01, Issue 02, December 2012.

[11]. P. Vinciya, Dr. A.Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining" IJARCSSE,Issue 5, 2016.

[12]. Shivnath Ghosh,Santanu Koley, "Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks" IJRITCC, Issue 2,292- 297,2014.

[13]. Zhihao Hong,Z. Kalbarczyk,R. K. Iyer,"A Data Driven Approach to Soil Moisture Collection and Prediction" IEEE-Xplore, Issue 2,292-297,2016.