International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

CROP YIELD PREDICTION USING MACHINE LEARNING AND SAGE MAKER

Manoj R

School of Computer Science and IT, Jain (Deemed-to-be University), Bangalore, Karnataka, INDIA

Manojstar205@gmail.com

Murugan R

School of Computer Science and IT, Jain (Deemed-to-be University), Bangalore, Karnataka, INDIA

muruganraam75@gmail.com

Abstract: Impact of climate change in India, many agricultural crops are badly affected by their performance over the past two decades. Predicting the harvest ahead of time can help policymakers and farmers to take appropriate marketing and storage measures. This project will help farmers to know the yield of their crop before sowing in the field and help them to make the right decisions. It tries to solve the problem by building a prototype of an interactive guessing system. Implementation of an easy-to-use web-based user interface and machine learning algorithm will be implemented. The results of the forecast will be made available to the farmer. Therefore, in such a type of data analysis in crop prediction, there are different methods or algorithms, and with the help of those strategies we can predict crop yields. Using a random forest algorithm. By analyzing all these problems and problems such as climate, temperature, humidity, rain, humidity, there is no suitable solution and technology to overcome the situation we are facing. In India, there are many ways to increase economic growth in the agricultural sector. Data mining also helps predict crop yields. In general, data mining is the process of analyzing data with different perspectives and summarizing it into important information. Random Forest is a popular and powerful machine learning algorithm capable of performing both subdivision and decontamination activities, which works by building dozens of Decision Trees during training and producing classroom output which is class mode (planning) or mean prediction (descent) of individual trees.

Keywords: Agriculture, Machine Learning, Cloud Computing, AWS sage maker, python IDE'S, Crop Prediction

1.Introduction

Agriculture is the backbone of the Indian economy. In India, agricultural yield primarily depends on weather conditions. Rice cultivation mainly depends on rainfall. Timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production of crops. Yield prediction is an important agricultural problem. In the past farmers used to predict their yield from previous year yield experiences. Thus, for this kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms we can predict crop yield. Random forest algorithm is used. Using all these algorithms and with the help of inter-relation between them, there are growing range of applications and the role of Big data analytics techniques in agriculture. Since the creation of new innovative technologies and techniques the agriculture field is slowly degrading. Due to these, abundant invention people are concentrated on cultivating artificial products that are hybrid products where there leads to an unhealthy life. Nowadays, modern people don't have awareness about the cultivation of the crops at the right time and at the right place. Because of these cultivating techniques the seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to insecurity of food. By analysing all these issues and problems like weather, temperature and several factors, there is no proper solution and technologies to overcome the situation faced by us. In India, there are several ways to increase the economic growth in the field of agriculture. There are multiple ways to increase and improve the crop yield and the quality of the crops. Data mining is also useful for predicting crop yield production. The main objectives are

- To use machine learning techniques to predict crop yield.
- To provide easy to use User Interface.
- To increase the accuracy of crop yield prediction. d. To analyse different climatic parameters (cloud cover, rainfall, temperature)

2.Literature Review

In [1] Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.This paper focuses on predicting the yield of the crop based on the existing Forest algorithm. Realdata of TamilNadu were used for building the models and the models were tested with samples. Random Forest Algorithm can be used for accurate crop yield prediction.

In [2] Random forests for global and regional crop yield prediction. PLoS ONE Journal. Our 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 regression (MLR).

In [3]. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and SoftwareEngineering (IJCSSE). In this paper, AdaNaive and AdaSVM are the proposed ensemble model used to project the crop production over atime period. Implementation done using AdaSVM and AdaNaive. AdaBoost increases efficiency of SVM and Naive Bayes algorithm.

In [4]. Analysis of Crop Yield Prediction by making Use Data Mining Methods. IJRET: The paper provided in International Journal of Research in Engineering and Technology. 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.

In [5]. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38),DOI:10.17485/ijst/2016/v9i38/95032, October 2016.From GPS based colour 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 thecrop yield for better enhancement and ratio of the yield can be increased during cultivation.

METHODOLOGY

Data is the most important part of any machine learning program. In order to implement the plan, we have decided to focus on Maharashtra Province in India. As the climate changed from place to place, data had to be obtained at regional level. Historical data on the crop and climate of a particular region were required to operate the system. This data is collected from various government websites. Information about the plants of each Maharashtra region was collected at www.data.gov.in and weather data was collected at www.imd.gov.in. The weather conditions that affect the crop the most are rainfall, temperature, cloud cover, humidity, and frequent wet days. Therefore, data on these weather conditions were collected at the monthly level.

3. Data Set Collection

In this section, we collect data from various sources and adjust data sets. And the provided database is used for statistics (descriptive and diagnostic). There are several sources of online summaries such as Data.gov.in and indiaastat.org. At least ten years a year plant abbreviations will be used. These databases are generally acceptable the behavior of the anarchic time series. Combined key once abbreviations required. Global Informal Forests and Regional Plants Yield Forecasts.

Karnataka	BELGAUM	1997	Kharif	Arhar/Tur	11268.00	1820.00
Karnataka	BELGAUM	1997	Kharif	Bajra	34917.00	5666.00
Karnataka	BELGAUM	1997	Kharif	Dry chillies	8266.00	12837.00
Karnataka	BELGAUM	1997	Kharif	Groundnut	75288.00	58174.00
Karnataka	BELGAUM	1997	Kharif	Horse-gram	10520.00	3578.00
Karnataka	BELGAUM	1997	Kharif	Jowar	52232.00	64259.00
Karnataka	BELGAUM	1997	Kharif	Maize	58549.00	181319.00
Karnataka	BELGAUM	1997	Kharif	Paddy	63734.00	137452.00
Karnataka	BELGAUM	1997	Kharif	Ragi	2501.00	2525.00
Karnataka	BELGAUM	1997	Kharif	Rice	63734.00	91677.00
Karnataka	BELGAUM	1997	Rabi	Dry chillies	579.00	324.00
Karnataka	BELGAUM	1997	Rabi	Groundnut	6733.00	5565.00
Karnataka	BELGAUM	1997	Rabi	Horse-gram	2833.00	964.00
Karnataka	BELGAUM	1997	Rabi	Jowar	140805.00	63996.00
Karnataka	BELGAUM	1997	Rabi	Maize	26432.00	50103.00
Karnataka	BELGAUM	1997	Rabi	Wheat	55130.00	41220.00
Karnataka	BELGAUM	1997	Summer	Jowar	338.00	834.00
Karnataka	BELGAUM	1997	Summer	Maize	3484.00	6721.00
Karnataka	BELGAUM	1997	Summer	Paddy	40.00	171.00
Karnataka	BELGAUM	1997	Summer	Rice	40.00	114.00
Karnataka	BELGAUM	1997	Whole Year	Coriander	215.00	26.00
Karnataka	BELGAUM	1997	Whole Year	Garlic	765.00	915.00
Karnataka	BELGAUM	1997	Whole Year	Sugarcane	125173.00	11177949.00
Karnataka	BELGAUM	1997	Whole Year	Turmeric	818.00	4496.00
Karnataka	BELGAUM	1998	Kharif	Arhar/Tur	8602.00	3645.00

Fig 3.1 Dataset Used

4. Data Classification

All databases are divided into 2 parts: for example, say, 75% of the data is used for model training and 25% of data is set aside for model testing.

5. Data Model training in SageMaker

is done on machine learning compute instances. When a user trains a model in Amazon SageMaker, he/ she creates a training job.

• Training jobs comprise of:

i. S3 bucket (within the compute instance): The URL of the Amazon S3 bucket where the training data is stored
ii. ii. AWS SageMaker on ML instance: Compute resources or Machine Learning compute instances
iii. iii. S3 bucket (outside the compute instance): The URL of the Amazon S3 bucket where the output will be stored
iv. iv. Inference code image: The path of AWS Elastic Container Registry path where the code data is saved

• The input data is fetched from the specified Amazon S3 bucket 16

• Once the training job is built, Amazon SageMaker launches the ML compute instances

• Then, it trains the model with the training code and dataset • SageMaker stores the output and model artifacts in the AWS S3 bucket • In case the training code fails, the helper code performs the remaining task

• The interference code consists of multiple linear sequence containers that process the request for inferences on data.

• EC2 container registry is a storage registry that helps users to save, monitor, and deploy container images

6. Validating a Model With Sage Maker

You can evaluate your model using offline or historical data:

i. Offline Testing Use historical data to send requests to the model through Jupyter notebook in Amazon SageMaker for evaluation.

ii. **Online Testing with Live Data** It deploys multiple models into the endpoint of Amazon SageMaker and directs live traffic to the model for validation.

iii. Validating Using a "Holdout Set" Here, a part of the data is set aside, which is called a "holdout set". Later, the model is trained with remaining input data and generalizes the data based on what it learned initially.

iv. K-fold Validation Here, the input data is split into two parts. One part is called k, which is the validation data for testing the model, and the other part is k − 1 which is used as training data. Now, based on the input data, the machine learning models evaluate the final output

7. Machine Learning Algorithms

Supervised learning: Supervised machine learning algorithms can apply what you have learned in the past to new data using labeled examples. After Enough training the system can provide the objectives of any new inputs. IN ORDER to modify the learning appropriately the learning algorithm can also differentiate its results by correct, targeted and error detection. Uncontrolled learning: By comparison, the unregulated machine learning algorithms are used when the information used for training is not labeled and unseparated. Unattended reading analyzes how systems can perform the task of defining a hidden structure from unlabeled data. To define hidden properties from non-labeled data the system does not detect the correct output, but scans the data and may draw predictions from the data sets.

Random Forest Classifier: Random Forest is a popular and powerful machine learning algorithm capable of performing both subdivision and decontamination activities, which works by building a number of deciduous trees during training and producing class results that are phases (descriptions) of individual trees. The more trees in the forest the more predictable the weather will be.

Decision Tree: A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes)

Polynomial Regression: In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x.

8. Architecture Diagram

The diagram mentioned below is for the proposed system. The system architecture is a conceptual model that describes the structure and behavior of multiple components and subsystems like multiple software applications, network devices, hardware, and even other machinery of a system.

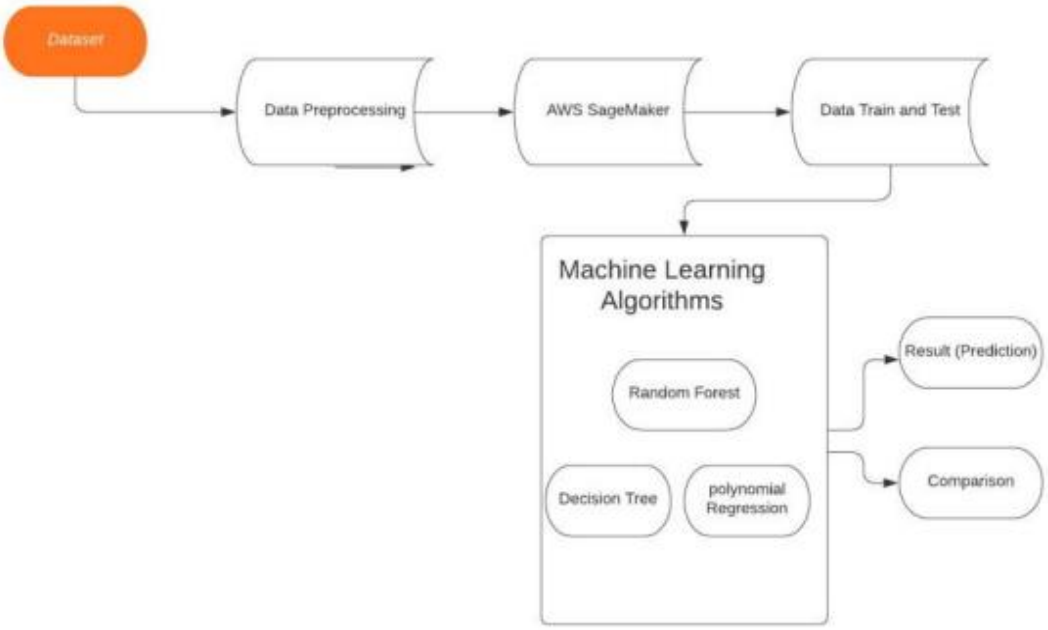

Fig 4.1 System Design

Firstly the Dataset is loaded into S3 bucket and then data is processed using the AWS Sage Maker and further the data is trained and tested using the machine learning algorithms like random forest, decision tree, and polynomial regression and then the result can be generated and can be compared with the remaining dataset for the accuracy.

9.Results

We have successfully developed the "Crop yield prediction using machine learning and sage maker, In the below diagram you can see the accuracy of the model for each for sunflower, maze, jowar crops

Fig 9.1 Crop Yield Prediction Result

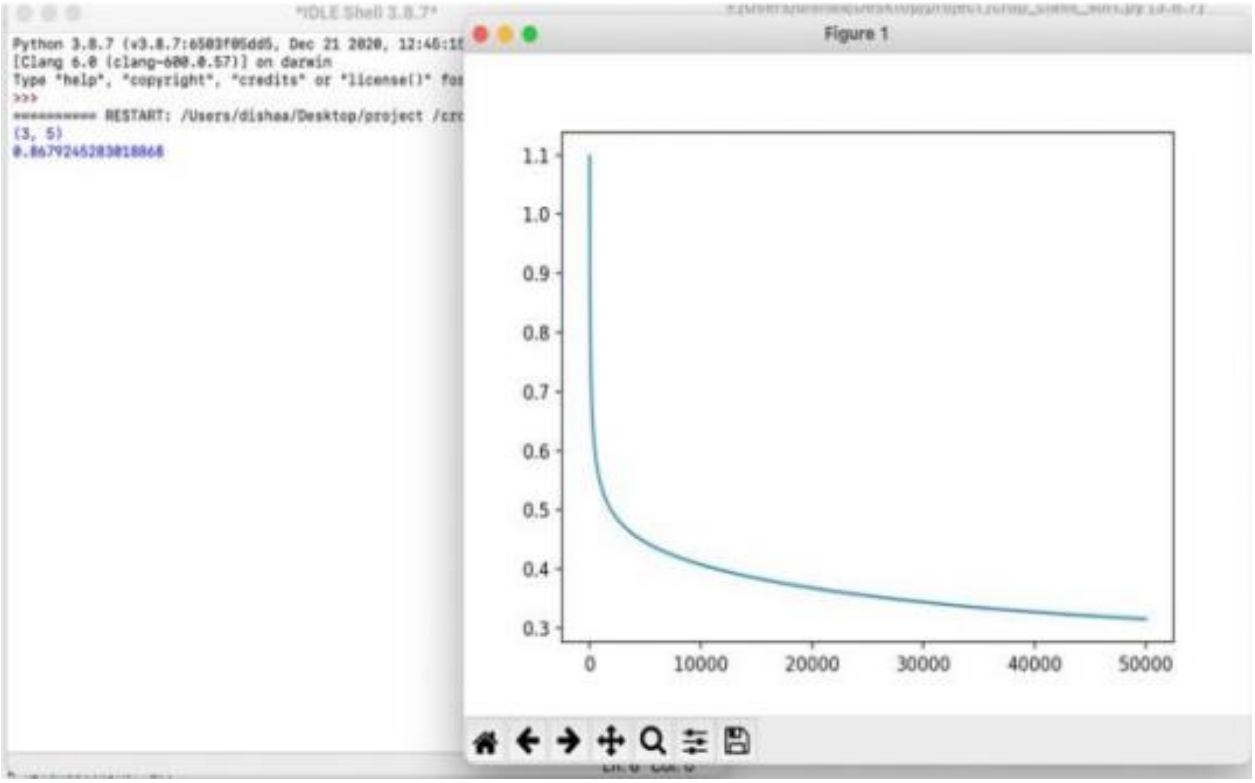

Fig 9.2 Crop Yield Prediction Graph

10. Future Enhancement

In the future, the project can be integrated with National Agriculture Department collecting the details from all over INDIA, Increase in the exact location based prediction of the users and the can be integrated with Private entities which helps the Customers for getting the better yield and good crop quality

11.Conclusion

This project is undertaken using machine learning and evaluates the performance by using Random forest, Polynomial Regression and Decision Tree algorithms. In our proposed model among all the three algorithm Random forest gives the better yield prediction as compared to other algorithms. Along with random forest, Polynomial Regression, Decision Tree model classify the output that shows improvements in dataset. So we analysed that proposed model has got more efficiency than the existing model for finding crop yield

References

[1] P.Priya, U.Muthaiah M.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2]. J.Jeong, J.Resop, N.Mueller and team.Random forests for global and regional crop yield prediction.PLoS ONE Journal.

[3].Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy.Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5]. Shweta K Shahane , Prajakta V Tawale.Prediction On Crop Cultivation. IInternational Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

[6]D Ramesh ,B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology.

[7]Subhadra Mishra,Debahuti Mishra, Gour Hari Santra. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016.

[8].Konstantinos G. Liakos,Patrizia Busato,Dimitrios Moshou, Simon Pearson ID,Dionysis Bochtis. Machine Learning in Agriculture. Lincoln Institute for Agri-food Technology (LIAT), University of Lincoln, Brayford Way, Brayford Pool,Lincoln LN6 7TS, UK, spearson@lincoln.ac.uk.

[9]. Baisali Ghosh. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Atur Centre, Gokhale Cross Road, Model Colony, Pune – 411016