iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43703

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Harvest Forecasting of Crop Yield Using Machine Learning

Mr. Ram Prashath R M.C.A., M.Phil.[1], Krishna Prasad K[2]
[1]*Assistant Professor,* [2]*Scholar, Department of MCA, Karpagam College of Engineering, Coimbatore, India*

Abstract: ***As an agriculturally dependent country, India's economic standing is totally and partially reliant on it. Agricultural yield is influenced by Causes include organic, economic, and seasonal factors. Calculation of this country's agricultural output is a major challenge. Taking into account the current population situation Recently, Nowadays, the people that grow these and similar items include Due to the suddenness of the production, it is extremely unstable. environmental factors such as weather and a lack of groundwater resources. The major goal is to collect data that can be used to make decisions. stored and analyzed for crop yield predictions For Machine learning techniques for agricultural yield prediction implemented. This aids farmers in selecting the best crops. appropriate crop In addition, this study tries to provide an enhancement in the world of agriculture, by improving crop production prediction accuracy. A statistical model is constructed using machine learning techniques and sufficient optimizations to produce accurate and precise decisions. The results of this research will assist farmers in selecting the best crops to cultivate based on characteristics such as season and available land, with the least amount of risk.***

Keywords: logistic regression, naïve bayes, random forest

I. INTRODUCTION

Agriculture has been the primary and preeminent activity of every culture and civilization throughout human history since its beginnings. It is not only a huge part of the rising economy, but it is also necessary for our survival. It is also a vital sector for the Indian economy and the future of humanity. It also employs a significant number of people. With the passage of time, the demand for production has grown enormously. People are exploiting technology incorrectly in order to create in large quantities. Every day, new hybrid kinds are developed. However, these variants lack the crucial nutrients found in normally grown crops.
These artificial methods degrade the soil. It all leads to more environmental damage. The majority of these artificial methods are used to avoid losses. However, when agricultural farmers have precise information on crop production, the loss is minimized. Machine learning is a rapidly expanding approach that is assisting every sector in making viable decisions to develop the most important of its applications. Most modern devices benefit from the analysis of models before to implementation. The fundamental idea is to use Machine Learning models to boost the throughput of the agriculture sector. Another aspect that influences prediction is the amount of knowledge imparted throughout the training period, as the number of parameters was relatively larger.

II. LITERATURE SURVEY

The agricultural sector plays an important role in the Indian economy. In India, 70% of households are dependent on agriculture. Agriculture contributed 17 percent to India's Gross Domestic Product in 2015-2016. However, agribusiness' commitment to Gross Value Added has been steadily declining. Because food is crucial for life and we rely mostly on agricultural harvests, farmers play a critical role. For the investigation of crop, the work used the Regression algorithm. In order to examine more than 362 datasets and provide organization, choice tree calculation and classification are used. For predicting the kind of soil, the preparation informational index is diverted into natural, inorganic, and land.
This framework keeps track of arrangements. precise and dependable The data to access the test informational collection was taken care of in. Back Propagation Network makes use of a shrouded layer that aids in the prediction of soil qualities. This results in great precision and performance superiority as compared to previous methodologies; however, the framework can become sluggish and irregularity in the yield can occur. Two relapse-managed AI techniques are used in the paper for work: In terms of soil quality expectations, SVM and RVM show viability. A clever remote technology detects soil wetness, and weather data is used. The remote control device has a 15% error rate and a precision of 95%. Regardless, it has not been tried in the current information. The investigation includes using back engendering calculations to verify soil fertility and plant nutrients. The yields are precise, allowing for the improvement of soil qualities. When compared to traditional tactics, this yields superior results. Framework, on the other hand, is moderately wasteful and inconsistent. Essentially, centers are using soil factors to improve crop yield. Sujatha et al. explain the rationale behind various crop yield prediction strategies. J48 is an example of a small information mining strategy. Artificial neural systems, arbitrary timberlands, bolster vector machines, and fake neural systems were all introduced. A framework for measuring crop development using atmospheric data and yield metrics has been

proposed. Kushwaha et al. provide the appropriateness of yield for various climatic features, as well as potential outcomes in improving harvest quality by using climate and illness-related informative indexes. They provided an inquiry, order, and prediction calculation in, which is useful in constructing a choice emotionally supporting network for developing accuracy.

III. METHODOLOGY

A. Data Cleaning

As a prelude to the pre-handling stage, we dropped such high cardinality factors during this method.

B. Pre-processing And Transformation

The objective variable was removed from the entire data set, and the straight out factor was replaced with a model framework with one-hot encoding. This is occasionally the case when specific calculations are required to process data in a limited framework group. When developing models, other quantifiable programming, such as R, automates this technique. I assigned a value of 0 to the information's missing attributes. I used the min-max standardization to scale the persistent factors, which shifts esteems onto a scale from 0 to 1 to square factors on multiple scales, greatly impacting the coefficients.

C. Data Partition

We divided the pre-processed data into preparation and test informational index. Data partitioning is a supervised machine learning classification that is used to separate pre-defined data classes. This is frequently used to divide data into numerous parts. Both training and testing datasets can be obtained using data partitioning. We have created a
k-NN classifier model with 10 neighbour classes and the Euclidean separation square factors on different scales heavily influencing the coefficients.

D. Data Evaluation

We currently determined the R squared data for both the preparation and test information by scoring the classifier on inconspicuous test information.

IV. CLASSIFIERS USED

Logistic Regression, Random Forest, and Nave Bayes were the machine learning classifiers utilized for accuracy comparison and prediction. The dataset was used to train these three classifiers.

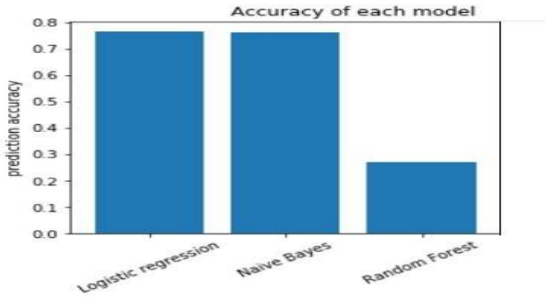


Fig.1. Accuracy of model

V. DATASET

	Month	Year	Rainfall	WPI
2	4	2012	47.5	101.2
3	5	2012	31.7	101.5
4	6	2012	117.8	102.5
5	7	2012	250.2	110.7
6	8	2012	262.4	115
7	9	2012	193.5	112.5
8	10	2012	58.7	114
9	11	2012	30.7	111.9
10	12	2012	11.7	110.8
11	1	2013	11.3	106.5
12	2	2013	40.1	106.1
13	3	2013	15.7	112
14	4	2013	30.4	112
15	5	2013	57.8	114.4
16	6	2013	219.8	115.9
17	7	2013	310	119.1
18	8	2013	254.7	123.6
19	9	2013	152.7	120.4
20	10	2013	129.4	118.7

Fig.2.Processed dataset sample

VI. BAR GRAPH OF CROPS

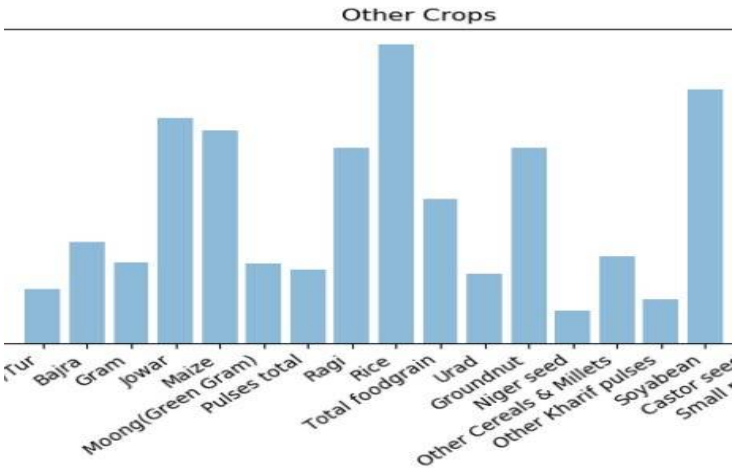


Fig.3.Bar graph

VII. CONCLUSION

Through the use of machine learning algorithms, this research focuses on crop forecast and yield computation. For the calculation of accuracy, several machine learning approaches were applied. The crop forecast for the given district was done using the Random Forest classifier. Implemented a system for crop prediction based on historical data. Farmers can use the proposed technique to help them decide which crop to plant in their fields. This work is used to find out more about the crop so that it may be used to harvest it efficiently and effectively.

This helps to boost the Indian economy by increasing crop output rates. In the future, ranch vehicles such as tractors could be connected to the internet, allowing ranchers to receive more information on yield gathering and potentially tainted harvests, allowing them to make better decisions. In addition, the most productive harvest may be found when the financial and expansion proportions are taken into account.

REFERENCES

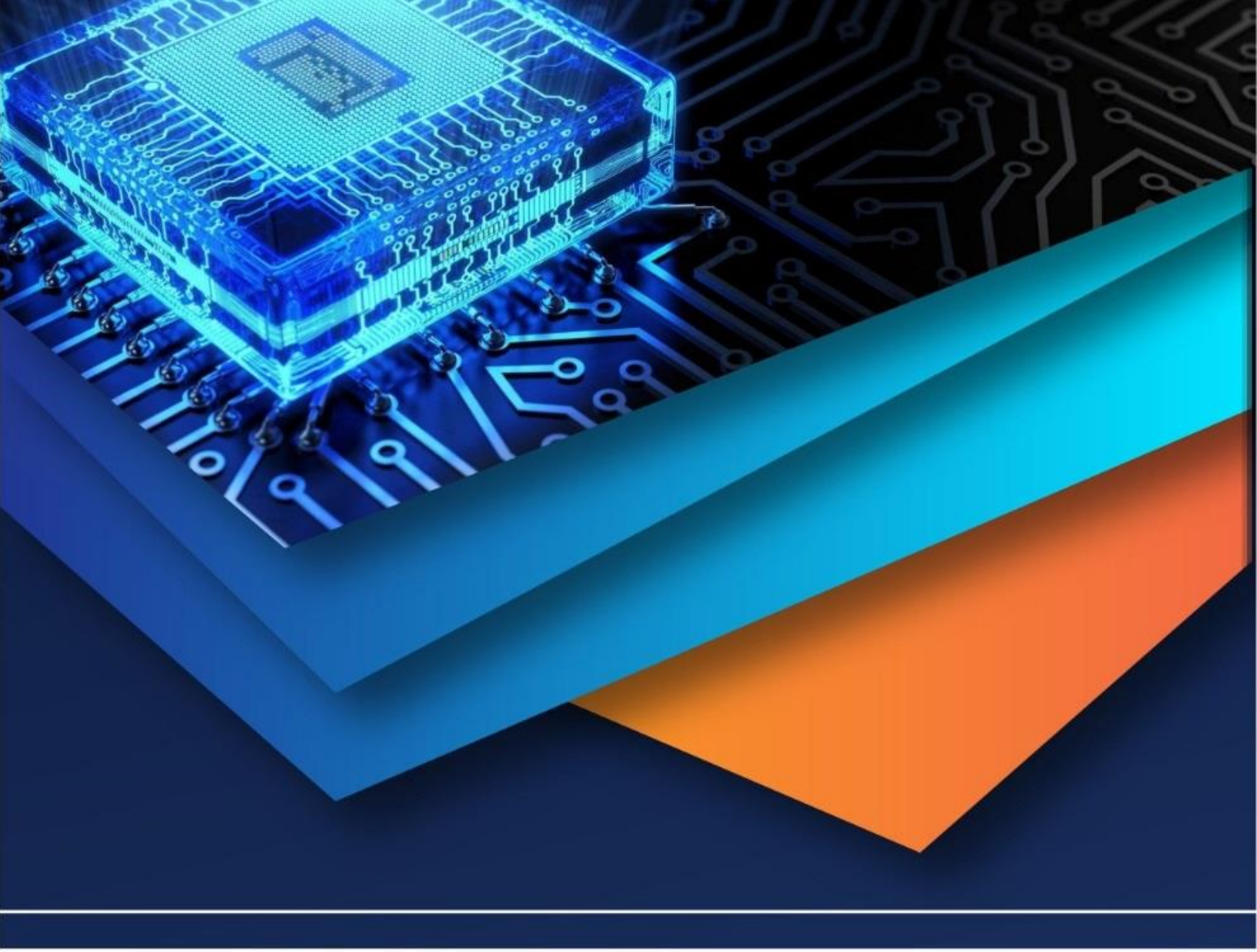
[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal "Crop Yield Prediction using ML Algorithms ",2019

[2] Gour Hari Santra, Debahuti Mishra and Subhadra Mishra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016

[3] Ramesh Medar,Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2019

[4] Sangeeta, Shruthi G, "Design And Implementation Of Crop Yield Prediction Model In Agriculture",2020

[5] data.gov.in



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY