International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 5, pp 297-299, May-2023

https://doi.org/10.47001/IRJIET/2023.705041

Crop Yield Prediction Using Machine Learning

[1]Kamalesh M, [2]Dr. Ragaventhiran J

[1]M. Tech. Artificial Intelligence, Department of CSE, School of Engineering, Presidency University, Bangalore, India

[2]Professor, School of Engineering, Presidency University, Bangalore, India

***Abstract* - Over 50% of India's population depends on agriculture for existence, making it the foundation of the country's economy. Variations in weather, climate, and other environmental factors are now a significant threat to the continued success of agriculture. The decision support tool for Crop Yield Prediction (CYP), which includes assisting decisions on which crops to plant and what to do during the growth season of the crops, is where machine learning (ML) plays a vital role. The goal of the current study is to conduct a systematic review that extracts and synthesises the CYP traits. In addition, a number of methodologies have been created to analyse agricultural yield prediction utilising artificial intelligence techniques. Reduction in relative error and lower crop yield prediction accuracy are the Neural Network's main drawbacks. Similar to this, supervised learning algorithms failed to recognise the nonlinear relationship between input and output variables, which presented a challenge during the selection, grading, or sorting of fruits. To establish an accurate and effective model for crop classification, including crop yield estimation based on weather, crop disease, classification of crops based on the growing phase, etc., numerous investigations were advised. This study examines various machine learning (ML) approaches applied to agricultural yield estimation and provides a thorough review of the strategies' accuracy.**

Keywords: Crop, Machine Learning, Crop Yield Prediction, CYP, Agriculture.

I. INTRODUCTION

The application of machine learning (ML) techniques may be seen in many different contexts, from predicting customer phone usage to examining customer behaviour in supermarkets. Agriculture has long utilised machine learning. One of the most challenging problems in the realm of horticulture is the prediction of crop production, and several models have undoubtedly been put up and evaluated up to this point. Due to the fact that various factors, including soil, climate, environment, compost use, and seed selection, have an impact on harvest yield, this problem calls for the use of a number of datasets. This indicates that determining crop yields is not a straightforward interaction but rather a series of intricate developments. The real yield can now be calculated with ease by yield expectation models, but a better yield projection yield is still required. Regression techniques are used to forecast the future, and descriptive models are used to interpret the data and explain what has occurred. ML studies provide a variety of difficulties when trying to build a high-performance predictive model. The correct algorithms must be chosen for the task, and both the underlying platforms and the algorithms themselves must be able to handle massive amounts of data. Despite being the principal activity of more than 50% of the population in India's rural areas, agriculture only contributes 17% of the nation's GDP, according to estimates from 2018. Using information that has been passed down through the ages, farmers grow crops today.

Since the traditional method, there has either been an excess or a deficiency of such yields without accumulating the genuine demand for cultivating a particular crop. This is the major obstacle facing Indian farmers. Farmers are killing themselves in greater numbers due to low crop prices, family pressure, and hefty development costs. Around the world, soil resources are being lost, water is being contaminated, and farmer incomes are gradually falling. This issue can be resolved by trying a new crop, but farmers find it challenging to predict the yield they will get from a specific crop. If they can forecast which crop would provide the highest price, they will profit more and help farmers. A recent innovation in agriculture called yield estimation has had a significant economic impact. To gather pertinent information, it makes use of satellite photos, extensive databases of soil, climate, and plant history characteristics.

We've selected to employ the top-performing regression models in order to assess their accuracy and determine which model can accurately forecast yield under specific weather and soil conditions. A dataset is a crucial part of a sophisticated system. We made the decision to introduce this method in Maharashtra. We obtained district-level data since climate factors depend on districts. This system required data on the weather, soil, and yield from previous years. Data on crops were compiled by district from the Indian agricultural website. Regression models are being used to forecast the yield, a continuous variable. The factors that affect crop production the most are rainfall, humidity, temperature, area, soil type, crop type, and season. So, for this approach, the aforementioned parameters are collected, district-by-district, throughout the last 18 years. The parameters were pre-

ISSN (online): 2581-3048
Volume 7, Issue 5, pp 297-299, May-2023
https://doi.org/10.47001/IRJIET/2023.705041

processed to fit the training data, and the objective was to identify the regression model that most accurately predicted crop yield. We utilised Python to develop the regression models, minimising the error between the real and projected values during model training. Different machine learning models that are effective at regression and produce the best outcomes have been used in this research. Metrics like R squared and mean absolute error are used to assess their performance. In India, crops are sown all year long. Depending on the season in which they are grown, crops are categorised. In India, the two principal growing seasons are Kharif and Rabi. The growing seasons for Kharif and Rabi are respectively from October to March and July to October. A single dataset was then created by combining the crop and climate data. As independent variables, historical crop information including the crop's district, name, season, and other climate factors that affect yield were employed. The crop's yield will act as the dependent variable.

The following are some of the difficulties we ran across while working on this project:

- Choosing an appropriate dataset and pre-processing the data makes the project more desired to achieve the best results.
- Less computing power must be used during regression model training.
- A higher error rate because to the districts' constantly changing environment.

II. RELEVANT WORK

Simple RNN and LSTM were used by Aruvansh et al. [1] for climate expectation, and the irregular timberland approach was used for yield forecasting from anticipated climate bounds. In order to analyse the yield of 15 consistently established crops, various AI models, such as XG help, KNN classifier, Logistic relapse, Stochastic Gradient Descent Classifier, and ANN, were also applied. However, crop yield was only estimated using temperature and precipitation, while crop.

III. DATASET DESCRIPTION

For the purpose of predicting crop yield, we have compiled a farm-related csv dataset. It offers several factors for each district, including temperature, precipitation, humidity, soil type, crop type, season, and field size. We have to encode those variables because some models can't take string values as inputs. This will make it easier for the system to comprehend the columns and accept them as input. Despite the fact that each item is assigned a distinct integer value based on order, label encoding provides a few advantages.

IV. ML ALGORITHM

Multiple linear regression, or MLR, is a mathematical technique for predicting an answer parameter's outcome by using a number of logical elements. It shows how a continuous dependent parameter interacts with a number of independent parameters. It is the most straightforward version and does not correct for choosing any weights. This could lead to the model giving a certain characteristic a lot of weight, which could lead to over fit. A linear regression cannot be performed unless certain conditions are met.

V. RESULTS

A laptop with an i5-9300U processor, a GTX 1050 graphics chip, and 8 GB of RAM was used to implement the machine learning models that were a crucial component of this research project. We divided the csv file containing our dataset into test and training sets. Regression models will be used because we will be predicting a continuous dependent variable using multiple independent factors. Gradient boosting, Elastic Net, Lasso, Multiple Linear Regression, Decision Tree Regression.

VI. CONCLUSION

We used a variety of regression models to implement crop yield prediction in our project, and we discovered that the feature engineering-based LSTM is the most effective model. It has the best accuracy (86.3%), the lowest mean absolute error, and the lowest root mean square error when compared to other models. Accuracy can also be increased with more data. We created this project from beginning to end so that farmers may use our website to directly decide which new crops to try out in a particular field that will deliver the most benefit.

REFERENCES

[1] Nigam A., Garg S., Agrawal A., Agrawal P. (2019) Crop yield prediction using machine learning algorithms in 2019.
[2] Predicting yield of the crop using machine learning algorithm Int j eng sci res technol (2018).
[3] Hemageetha N. (2016) A survey on application of data mining techniques to analyze the soil for agricultural purpose.
[4] Sujatha R., Isakki P. (2016) A study on crop yield forecasting using classification techniques. In International.
[5] A.K. Kushwaha et al.Crop yield prediction using Agro Algorithm in Hadoop Int J Comput Sci Inf Technol Secur (IJCSITS) (2015).

ISSN (online): 2581-3048
Volume 7, Issue 5, pp 297-299, May-2023
https://doi.org/10.47001/IRJIET/2023.705041

[6] G.N. Fathima et al.Agriculture crop pattern using data mining techniques ,Int J Adv Res Comput Sci Softw Eng (2014).

Citation of this Article:

Kamalesh M, Dr. Ragaventhiran J, "Crop Yield Prediction Using Machine Learning" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 5, pp 297-299, May 2023. https://doi.org/10.47001/IRJIET/2023.705041

****

