IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Crop Yield Prediction Using Machine Learning

T. Cyril [1], P.V. Archana [2], G.D. Vignesh[3]

[1]Department of Electronics and Communication Engineering, St. Joseph's College of Engineering, TN, India,

[2]Department of Electronics and Communication Engineering, St. Joseph's College of Engineering, TN, India,

[3]Department of Electronics and Communication Engineering, St. Joseph's College of Engineering, TN, India,

Abstract— This study introduces a cutting-edge Crop Yield System utilizing Machine Learning, revolutionizing agricultural practices. Employing advanced algorithms, the system accurately predicts crop yields based on historical and real-time data, optimizing resource allocation and decision-making in farming. Our objective is to enhance precision agriculture by creating a reliable and adaptable model that considers diverse factors affecting crop yield variability. Integrating sensor data and satellite imagery, the system offers insights into optimal planting strategies and irrigation scheduling for sustainable farming. Continuous refinement through collaboration with farmers and stakeholders ensures practical applicability, while a user-friendly interface facilitates widespread adoption. This innovation marks a significant step towards data-driven, efficient, and sustainable crop management in modern agriculture.

Keywords—: Convolutional neural networks (CNN), Machine Learning (ML), Artificial Neural Networks (ANN), Application Programming Interfaces (API), Extreme Gradient Boosting (XGB) Support Vector Machine (SVM)

I. INTRODUCTION

In human history, agriculture has been the most important and important activity in all cultures and worlds since creation. This is not only an important part of economic expansion, but is necessary for our survival [1]. An economic activity highly dependent on the climate is agriculture. This indicates that rainfed agriculture, also known as seasonal agriculture, depends on weather conditions. Climate change affects agricultural yields and hunger and food insecurity. [2]. Crop yield forecasting is an important part of agriculture that helps farmers make decisions about their crops. This involves predicting the number

of crops that will be produced in a given area based on variables that include crop management practices, climate and soil type [3].

Predicting crop yields is one of the most difficult challenges. Predicting crop yields is one of the most difficult tasks in agriculture. It plays an important role in decision-making at the international, regional and field level [4]. Crop yield predictions based on soil, climate, environment and crop parameters [5]. Precision agriculture focuses on monitoring (intelligence technologies), management information systems, variable rate technologies and response to inter- and intra-variability of cropping systems. The benefits of sustainable agriculture include increased crop yields and product quality while reducing environmental impact.

Machine learning is a branch of artificial intelligence that allows computers to learn from data without being heavily programmed. This is great for product forecasting because you can identify patterns and relationships in large amounts of data and make predictions based on those relationships. Since the current models are in the understanding of defects, we use SVM to classify the product data based on the texture, shape and colour of the dead surface samples [6]. Applied methods use CNNs to improve crop yield predictions and reduce correlation errors [7].

Similarly, previous models used backpropagation neural networks (BPNNs) and time series models. However, the small number of samples used for prediction resulted in poor model performance [8], [9]. Implementing machine learning for crop yield prediction requires a large amount of crop yield data. This data includes product information, including product type, location and planting date. Once the machine learning algorithm is trained, it can be used to make predictions. Product forecasting in machine learning involves using data and predictive models to predict products and trends. It plays an important role in modern agriculture by helping farmers and agronomists make decisions about crop cultivation, management and harvesting [10].

II. PROPOSED SYSTEM

A crop advisory system can be designed using Scikit-Learn to help farmers make data-driven decisions about which crops to plant based on various conditions such as soil type, climate, and previous crop yield. Scikit-Learn is a popular machine learning library in Python that provides a wide range of tools to classify and tell farmers which crops will be most useful in a region and increase crop yield. The proposed model is crop selection based on economic and environmental factors, and the benefit of increasing crop yields will later help meet the country's increased food supply demand. Predict product results by studying factors such as state, region, territory, and weather. This system also helps determine the best time to apply fertilizers. Users provide state, region, season, yield, and acres as input to the output. Users provide state, region, time, and region to enter product recommendations. Depending on your needs, the model will indicate the performance of a specific product. The model recommends the most effective crops and indicates the right times for fertilizer application. The main goal is to grow more varieties over time. The Crop Recommendation System (CRS) utilizing ML techniques is designed to provide accurate and personalized

crop recommendations to farmers based on data analysis and predictive modeling. By harnessing the power of ML algorithms, the CRS aims to improve crop selection decisions and optimize agricultural productivity.

Data Processing: This part of the report loads, cleans, and refines the data to prepare the dataset for analysis. Be sure to follow the steps in your document and accurately report your cleaning decisions.

Data Collection: Training and test sets are created from data sets collected to predict the given data. A ratio of 7:3 is used to split the training and test sets. The training set is based on data models developed using Random Forest, Naive Bayes, Decision Trees and Logistic Regression. The test set prediction is done based on the accuracy of the test results.

Preprocessing: Missing values in the collected data will result in inaccurate results. Data preprocessing is necessary to improve algorithm efficiency and achieve better results.

In addition to removing the backs, the index changes must be completed.

Building a classification model: The following factors make the forecast of crop production an effective high-precision forecasting model: For a classification problem, it produces excellent results.

• It is effective in preprocessing a combination of outliers, irrelevant variables and discrete, continuous, and categorical data.

• It produces an estimated error that is fairly easy to change and has been shown to be unbiased in many tests.

Building a predictive model: Machine learning requires the collection of large amounts of historical data. Sufficient raw and historical data are available for data collection. Data preprocessing is necessary before raw data can be used directly. Any method using this template is then used to preprocess the form. This model has been trained and tested to ensure accurate predictions with few errors.

Construction of a Predictive Model: Machine learning needs data gathering have lot of past data. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data cannot be used directly. It is used to pre-process then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors.

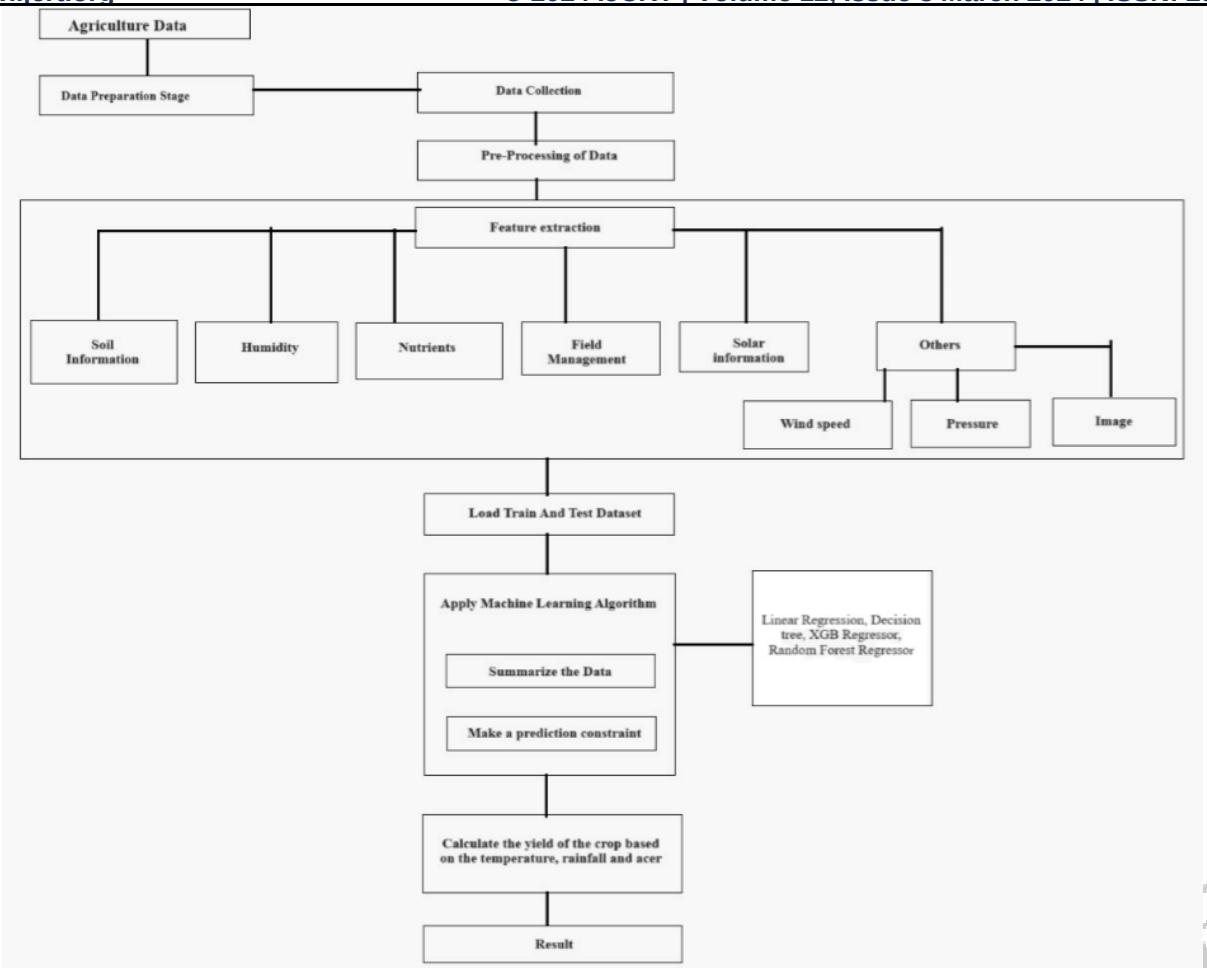

Fig 1. System Architecture

III. WORKING PRINCIPLE

LIST OF MODULES

- Pre-processing of the data
- Data analysis through visualization
- Decision tree regression algorithm implementation
- Linear regression algorithm implementation
- XGB regression algorithm implementation
- Random forest regression algorithm implementation
- Deployment Using Flask Framework

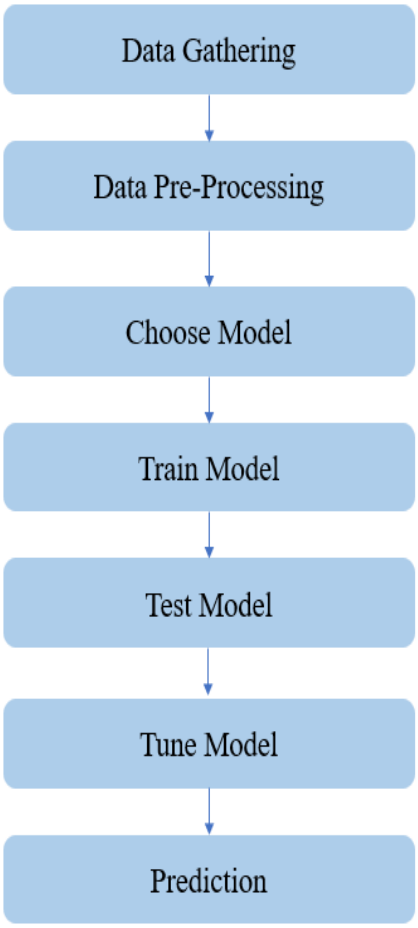

Fig 2. Work Flow Diagram

DATA PRE-PROCESSING

The workflow begins with data collection. This step involves gathering relevant data from various sources, which may include databases, external datasets, APIs, or sensors. Data collection may also encompass the annotation of data, such as labelling images or categorizing text.
After data collection, the raw data often needs to be cleaned and pre-processed. This step includes handling missing values, removing duplicates, and addressing outliers.
Python's Pandas module **is** a tool for many data cleaning tasks. You can clean your data faster by focusing on missing values, which can be the most important data cleaning tasks. Time is better spent exploring an d modeling your data than cleaning it.
Identifying variables using univariate, binomial**,**
and multi-variate analysis:

- import libraries for access and functional needs and analyse the given datasets
- General properties of examining the given dataset
- Display the given dataset in the form of data frame
- Displaying of columns
- format of the data frame
- Description of data frame
- Checking data type and information about dataset
- Checking for duplicate data
- Check for missing values in a data frame
- Check for unique values in a data frame
- Check for count values in a data frame
- Rename and drop the given data frame
- To specify the type of values

- To create additional columns

DATA VISUALISATION

Data visualization is the process of using visual elements like charts, graphs, or maps to represent data. It translates complex, high-volume, or numerical data into a visual representation that is easier to process. Data visualization tools improve and automate the visual communication process for accuracy and detail.

ALGORITHM IMPLEMENTATION

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple algorithms in python with sci-kit learn. A way to do this is to use different visualisation methods to show average accuracy, variance, and other properties of the distribution of model accuracies.

XGB Regressor:

The XG Boost (Extreme Gradient Boosting) Regressor is a powerful machine learning algorithm employed in the CPY. This algorithm belongs to the ensemble learning family and combines the predictions from multiple decision tree models to enhance accuracy and predictive performance. XGB Regressor is adept at handling complex relationships within agricultural datasets, providing nuanced insights into the factors influencing crop yields.

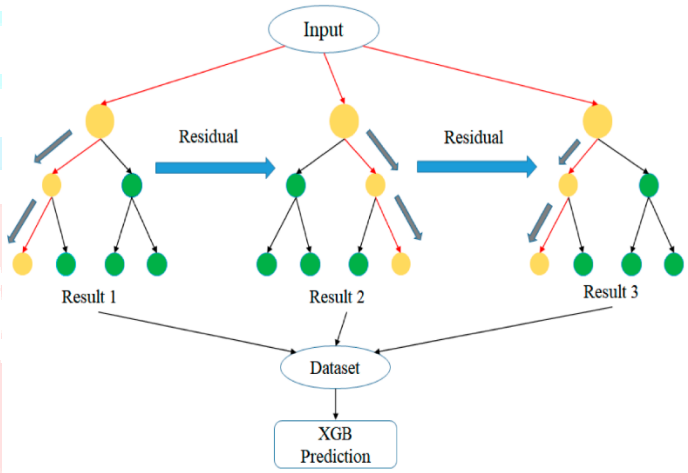

Fig 3. XGB Regression

IV. RESULT

Flask (Web Frame Work)

Python has an API called Flask framework that lets us create web apps.

Assemble your project: To get started, make a new Flask framework project. This will generate a new project directory with all the necessary files in it.

Define your models: Models are the objects that represent your data. For example, you might have a model for a user, a post, or a product. You can define your models in the `models.py` file in your app directory.

Create your views: Views are the functions that handle requests from users. For example, you might have a view that displays a list of posts, or a view that allows users to create a new post. You can create your views in the views.py file in your app directory.

Configure your URLs: URLs are the addresses that users use to access your website. You need to configure your URLs so that Django knows which view to call when a user visits a particular URL. You can configure your URLs in the urls.py file in your project directory. You can run your server.

Deploy your website: Once you are happy with your website, you can deploy it to a production server. There are many ways to deploy a Flask website, so you will need to choose a method that works for you.

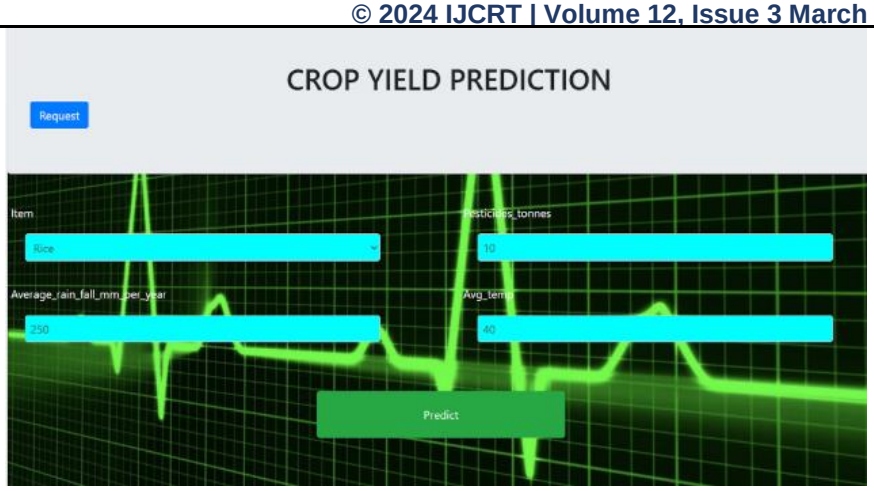

Fig 4. Factors to predict yield

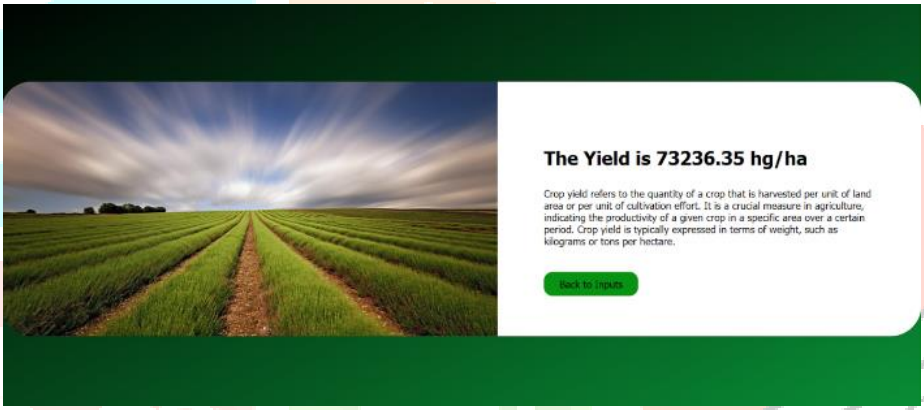

Fig 5. Yield predicted

The success of this case study highlights the potential of Machine learning technology for accurately predicting the crop yield in a particular environment. Further research can explore the integration of this system with existing agriculture management frameworks and investigate the impact on other factors that focusses on agricultural improvement.

V. CONCLUSION

The current study explored several scenarios that mostly depend on the availability of data, and each study will look at crop yield prediction using machine learning methods that are distinct from the features. The geological position, scale, and crop features were taken into consideration while choosing the features, and this process was mostly influenced by the availability of the data collection. However, using more features did not necessarily result in better outcomes. As a result, testing was done to identify the few highest-performing features that were also used in the research.

Our machine learning-based crop yield system demonstrates its potential to revolutionize modern agriculture. By harnessing advanced algorithms, we can accurately predict and optimize crop yields, empowering farmers with data-driven insights for sustainable and efficient farming practices. This novel strategy has the potential to address issues with global food security and guarantee a more resilient and fruitful agricultural future.

Further research should focus on critical possibilities, such as the need for additional-explicit therapy in the first instance due to the delay in border topographical areas. Afterwards, a nonparametric section of the model is developed using a machine learning technique. Lastly, features from deterministic crop models are utilized to provide optimal statistical CO2 fertilization.

VI. REFERENCES

[1] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, "Prediction of crop yield using machine learning," Int. Res. J. Eng. Technology, vol. 5, 2018.

[2] F. H. Tseng, H. H. Cho, and H. T. Wu, "Applying big data for intelligent agriculture-based crop selection analysis," IEEE Access, vol. 7, pp. 116965-116974, 2019.

[3] A. Suresh, N. Manjunathan, P. Rajesh, and E. Thangadurai, "Crop Yield Prediction Using Linear Support Vector Machine," European Journal of Molecular & Clinical Medicine, vol. 7, no. 6, pp. 2189-2195, 2020.

[4] M. Alagurajan, and C. Vijayakumaran, "ML Methods for Crop Yield Prediction and Estimation: An Exploration," International Journal of Engineering and Advanced Technology, vol. 9 no. 3, 2020

[5] P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Prediction of Crop Yield Using SVM Approach with the Facility of E-MART System" Easy chair 2020.

[6] S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthi veena, "A microcontroller-based machine vision approach for tomato grading and sorting using SVM classifier," Microprocessors and Microsystems, vol. 76, pp.103090, 2020

[7] P. Tiwari, and P. Shukla, "Crop yield prediction by modified convolutional neural network and geographical indexes," International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[8] P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1092-1096, 2020.

[9] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," Ictact journal on soft computing, vol. 11, no. 01, 2020.

[10] A. A. Alif, I. F. Shukanya, and T. N. Afee, "Crop prediction based on geographical and climatic data using machine learning and deep learning", Doctoral dissertation, BRAC University) 2018.

[11] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep learning-based detector for real-time tomato plant diseases and pests' recognition," Sensors, vol. 17, no. 9, pp. 2022, 2017.

[12] J. Sun, L, Di, Z. Sun, Y. Shen, and Z. Lai, “County-level soybean yield prediction using deep CNN-LSTM model,” Sensors, vol. 19, no. 20, pp. 4363, 2019.