International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 309-312
Published Online June 2023 in IJEAST (http://www.ijeast.com)

YIELD ENHANCER USING MACHINE LEARNING

Shruti Gupta, Yash Kumar Arora, Akshay Singh, Archita Verma
Student,
HMR Institute of Technology and Management,
New Delhi, India

Rekha Kumari
Assistant Professor,
HMR Institute of Technology and Management,
New Delhi, India

Abstract: **India a country with over one billion people stands at the second most populated country in the world. Indian economy is largely dependent on agriculture with around 50 percent of the country's GDP being supplemented by the primary sector. A large percentage of our population is involved in agricultural activities. But unfortunately struck at the old school methods for farming along with the rapid changes in climatic conditions a lot present farmers are facing issues their crop yield. This research is one step to battle with the current issues. This will helps the farmer to decide which crop to grow so that they can achieve maximum yield. Our system uses various machine learning algorithms like Decision tree, Gaussian Naive Bayes, Support Vector Machine (SVM), etc. The user is request input simple parameters like temperature, humidity, ph in order to get an answer.**

Keywords - **Machine Learning, Crop prediction, Decision tree, Logistic Regression, Decision Tree, XGBoost, Naïve Bayes, Random Forest, Fertilizer prediction, Crop recommendation;**

I. INTRODUCTION

India being an agrarian economy and supporting the livelihoods of millions of people, it is imperative that more and more farmers are supported through various measures to help them increase their yields and at the same time help improve their financial condition through technological advancements in farming. Traditionally farmers used to follow orthodox techniques to decide what crop to grow. These included farmers' knowledge about the given area or trending crops in the local areas. Owing to changing weather conditions and other phenomena like global warming sweeping the globe there are lot of fluctuations in crop yields and food availability is affected as a result. Due to the farmers' need for knowledge about soil nutrient content such as nitrogen, phosphorus, and potassium, they are facing challenges in deciding which crop would be best for maximum yields. Thus, Accurate prediction of crop yield and fertilizer requirements is critical for sustainable agriculture and food security.

Thus, Indian farmers must take a special interest in efficient and precise farming. For all these problems there is a very effective and simple solution which involves using machine learning. Machine learning helps us in predicting various measures that effect crop yields such as atmospheric conditions and the most suitable soil types. It is a game changer for agriculture as it can help farmers by comparing and contrasting various correlations and patterns and also help in increasing knowledge through various datasets. It focuses on learning from previous data and has the potential to augment the yield of crops. Historical data and various datasets are used to train the models.

This paper aims to create a user-friendly interface for farmers where the developed system will suggest the best suitable crop for a particular land. Based on atmospheric and weather parameters and soil content such as humidity, temperature, rainfall, ph, and amount of nitrogen, phosphorus, and potassium in soil. This model allows the farmers to input data in a user-friendly manner. All of the input data is then applied to predictive machine learning algorithms like Decision tree, Gaussian Naive Bayes, support vector machine, logistics regression, random forest and xgboost, with help of these algorithms the model process the data and identifies patterns based on which it recommends the most suitable crop to grow in that particular area.

II. LITERATURE REVIEW

This section focuses on the previous works and researches undertaken in the field of crop prediction and its related topics [1]. Majority of the research work has been done on predicting the yields of crops. Researchers B Vishnu and D Ramesh did research work based on rainfall dataset that focused on predicting yield. They researched on different

techniques and algorithms applied and used including Multiple Linear Regression (MLR). MLR involves finding a linear relationship when we have more than one independent variable and a single dependent variable. This research work also focuses on various techniques of Data mining and predicting yield mainly based on datasets of rainfall. It offers a MLR and K means comparison analysis as well [2]. Researchers Prof. (Dr.) Yusuf Mulge and Jyotshna Solanki also did research work on the different data mining techniques focused on agriculture. Disease detection, problem prediction, pesticide optimization are various topics on which research has been carried out. Multiple applications and Data mining are used to search and filter out important and relevant data included information that is linked from the vast databases and repositories of available information [3].

III. PROPOSED SYSTEM

The Proposed System will recommend and predict the most optimal crop to be sown and the fertilizers that the soil is deficient in based on parameters like temperature, humidity, soil pH, rainfall, moisture content etc. to give the best results to the farmers by enhancing their yields [4].

3.1 Data Collection

The data for our project was collected from various sources including but not limited to the Ministry of Agriculture and Farmers Welfare (Area, Production, Yield data), Development Department, Agriculture unit, Government of NCT Delhi (Fertilizer data), Kaggle etc. We have tried to diversify our data sources in order to get the best possible coverage of various parameters like temperature, pH, rainfall, moisture and areas sown to make our project more efficient and effective on a large scale.

3.2 Data Pre processing

The process of modifying raw data into a form that analysts and data scientists can use in machine learning algorithms to find insights or forecast outcomes is called Data preprocessing. In this project, the data processing method is to find missing values. Getting every data point for every record in a dataset is tough. Empty cells, values like null or a specific character, such as a question mark, might all indicate that data is missing. The dataset used in the project had few missing values and misspelled data which were removed after manual replacement of data fragments in the actual data.

3.3 Machine Learning Algorithms Used

3.3.1 Decision Tree

Decision Tree involves and is based on supervised learning technique which is used for problems involving paths where each set of decisions end up leading to a particular class. It is used in both regression and classification. A sequence of questions is asked by taking an instance from the training set. The non-terminal nodes such as root and internal nodes have decision attributes. The splitting continues generating sub-trees until it reaches a leaf node which determines class labels for that instance. It uses recursive partitioning which divides the tree recursively. With high accuracy, decision tree are capable of handling high-dimensional data. It closely parallels human-level thinking and involves a flowchart diagram-style representation. The accuracy achieved for it was: 90%

3.3.2 Random Forest

Random forests also known as **random decision forests.** Itis used for regression and classification and uses ensemble learning technique. During training of data, it constructs various decision trees. The individual tree's mean or average prediction is returned for regression tasks and the random forest's output is the class selected by most trees for classification tasks. The habit of Decision trees of overfitting to their training set is corrected through random decision forests. The accuracy of random forest is way lower than other tree algorithms including gradient boosted trees. However, their performance is better than decision trees and is affected by data characteristics. The accuracy achieved for it was: 99%

3.3.3 Naïve Bayes

Naive Bayes based on Bayes Theorem is a set of classification algorithms. It is a family of algorithms not just a single algorithm where all the algorithms operate on a common principle, i.e. every pair of feature independent of every other feature being classified. The accuracy achieved for it was: 99%

3.3.4 Logistic Regression

Logistic regression is a supervised learning algorithm based on classification. Classification problems involve output or target variable which is y and it can take only those values which are discreet from the set of input variables which is x. Logistic regression is largely categorized as a regression model. The model is used to predict probability and builds a regression Model for a data entry that belongs to the category which is numbered as "1". Just how a linear function is used in Linear regression to assumes linearity in data, Logistic regression uses sigmoid function and models the data on it. The accuracy achieved for it was: 95%

3.3.5 XGBoost

XGBoost uses Gradient Boosted decision tree's implementation. XGBoost models are dominant in many problems that aim to solve real life problems.

This algorithm uses decision trees in a sequential form. XGBoost is largely based on utilisation of weights. All independent variables are assigned some weights to predict results by feeding the weights into decision trees. The wrong predictions involve increasing the weight of variables

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 309-312
Published Online June 2023 in IJEAST (http://www.ijeast.com)

and feeding the increased weights to the second decision tree used. The predictions and classifiers are ensembled together to make a stronger model rather than just using them individually. It makes for a more precise model. XGBoost can work on classification as well as regression and also help solve user-defined prediction problems. The accuracy achieved for it was: 99%

IV. CROP PREDICTION

The Crop Prediction Model will help the farmers know the best crop to grow based on the nutrients present in the soil and other conditions like temperature, humidity, pH, rainfall etc. It also takes into consideration the area where a particular crop is grown to give the best possible yield to the farmer.

V. FERTILIZER PREDICTION

The Fertilizer Prediction model will predict the most suitable fertilizer to the farmer on the basis of crops grown, and the nutrients that the soil is deficient in. This will help to enhance the farmer's yield and also help in replenishing the soil will the right kind of fertilizers.

VI. CROP AND FERTILIZER RECOMMENDATION

As per the Department of Agriculture & Farmers Welfare (DAFW) the 20.2 percent is the share of agriculture and allied sector in GVA in year 2022. Which stated that India is an agro based country and maximum GDP is dependent on this. In this situation crop yield plays a vital role in the economic upswing of the nation. The system will suggest the most suited crop for cultivation based on the amount of anticipated rainfall, the composition of the soil, and meteorological data. This approach also displays the amount of seed needed to cultivate a recommended crop in kilograms per acre and offers information on the necessary fertilizers, such as phosphorus, nitrogen and potassium, NPK in kilograms per hectare. In the front end, there are various input parameters which are supposed to be filled by the farmers. A machine learning model is constructed at the back end taking into account all the data from datasets, the location, and the meteorological information. The Naive-Bayes algorithm is employed to forecast the results. Based on the appropriateness of the soil, crops are projected. The user can select any crop from the list of projected crops. Additionally, information on each expected crop is included. The user interface (UI) is highly interactive and responsive. The external crop datasets are loaded as part of the crop forecast procedure. After reading the dataset, pre-processing will go through different phases as described in the Data Pre-processing section. Train the models with a decision tree classifier using the pre-processed data. We take into account a number of variables for crop forecast, including temperature, humidity, soil PH, and anticipated rainfall. These are the input parameters for a system, and they can be manually. A list will be appended with the predicted rainfall and input parameter values. Using data from the list, the decision tree algorithm will forecast the crop.

VII. EXPERIMENTAL OUTCOME

The developed software will support any browser whether it is Safari, Chrome, Brave or any other. Users can also use our website on their mobile devices as we have made it responsive and user-friendly. The users must have their own smartphones, or laptop and a stable internet connectivity.

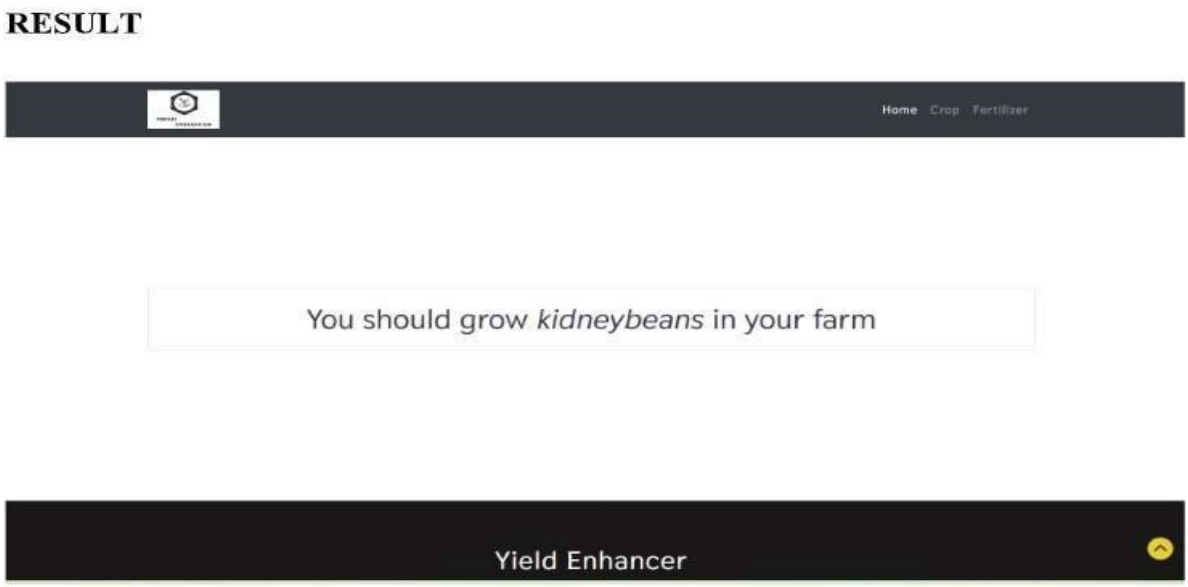

Figure 1: Output

VIII. CONCLUSION

We the GenZ kids are able to make farmers adopt the new upcoming technologies which will be very useful for them. In this research, we gathered the datasets from multiple places like government websites, third party sources and many others. We have created a user-friendly interface using GUI that will forecast which crop would be most suitable for a certain part of land and that will provide the necessary fertilizer information to increase yield. We can help the farmers to maximize their yields as per the few parameters they provide.

IX. FUTURE SCOPE

Currently we have a data constraint, and in future we want this data to provide accurate information of any location we search for. We will be trying to incorporate image processing techniques for agricultural disease prediction by gathering photographs of leaves, crops, and other plant parts. Gather enough information from all the Indian states and districts to enable farmers all around the nation to utilize the program. Additionally, raise model effectiveness. There shouldn't be a language barrier, and the software should be adaptable to everyone regardless of our country's diversity.

X. REFERENCE

[1]. Vijay Hedge S, Yasvanth C V, S Chandra Kiran "Crop Yield Prediction".
[2]. B Vishnu, D Ramesh "Analysis of crop yield prediction using Data mining techniques".
[3]. Prof. (Dr.) Yusuf Mulge, Jyotshna Solanki "Crop yield prediction in agriculture using Datamining predictive analytic techniques".
[4]. Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R "Crop Prediction using Machine Learning Approaches".