Crop yield prediction and Fertilizer Recommendation using Voting Based Ensemble Classifier

K. Pragathi
PG Scholar, Dept. of computer science and System Engineering(A), Andhra University College of Engineering ,Vishakhapatnam, Andhra Pradesh

Abstract **- India being an agriculture country, its economy predominantly depends on agriculture yield growth and agroindustry products. Data Mining is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect and what is the crop that is suitable for the land. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms like SVM, Random Forest, KNN and voting classifier for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with the proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue. Thus, the proposed system takes the data regarding the quality of soil and the weather-related information as an input. The quality of the soil such as Nitrogen, Phosphorous, Potassium and Ph value. Weather related information like Rainfall, Temperature and Humidity to predict the better crop. In our project we are taking the datasets from Kaggle website.**

Index Terms **- Crop, SVM, KNN, Random Forest, Nitrogen, Phosphorous, CNN**

1.INTRODUCTION

One of the most essential occupations in our country is agriculture. It is the country's most diverse economic sector and plays a critical role in its overall development. To meet the demands of the country's 1.2 billion inhabitants, agriculture takes up around 60% of the country's land. As a result, agricultural modernisation is critical, and it will lead to profit for our country's farmers. The act of studying data collections in order to derive conclusions about the information they contain, increasingly with the use of specialised tools and software, is known as data analysis (DA). Previously, yield prediction was done by taking into account the farmer's expertise on a specific field and crop Farmers, on the other hand, are forced to produce more and more crops as the weather changes swiftly from day to day. Given the existing scenario, many of them lack sufficient knowledge of the new crops and are unaware of the benefits they receive from producing them. Understanding and anticipating crop performance in a variety of environmental circumstances can also boost agricultural output. As a result, the suggested system uses data on soil quality as well as weather-related information as input. The soil's quality, such as nitrogen, phosphorus, potassium, and pH. Rainfall, temperature, and humidity are examples of weather-related information. In this project, we will Crop forecast is a common issue that arises. A farmer was curious about how much output he should expect during the rising season. Previously, this yield estimate was based on the farmer's long-term expertise with specific yields, crops, and meteorological circumstances. Instead of worrying about crop forecast, farmers go straight for yield prediction with the current technique. Unless the correct crop is forecasted, the yield will be better, and pesticides, environmental and climatic parameters related to the crop will not be considered using existing methods. One of the most important conditions for agricultural progress is to promote and soothe agricultural production at a faster rate. Any crop's production will guide the way either through domain interest or yield

enhancement, or both. In India, the only way to broaden the district under any crop is to re-establish to boost cropping strength or to replace the crop. As a result, changes in crop productivity continue to be a source of concern and distress in the area. As a result, in order to solve the problem, good crop prediction techniques must be tried.

2 LITERATURE SURVEY

VIRENDRA PANPATIL ET It had done a tremendous amount of work for Indian ranchers by creating a framework for producing productive yield proposals. They developed a framework that included classifier models such as Decision Tree Classifier, KNN, and Naive Bayes Classifier. The proposed framework can be used to determine the ideal planting season, plant development, and plant harvesting. They used a unique classifier to get higher exactness, for example: decision trees have poorer precision when datasets include more variations, while Nave Bayes has superior exactness than choice trees for such datasets. The most advantageous position of framework is that it can be really versatile all things considered/be utile

MAYANK ET It is anticipated that this paper will construct an extemporised framework for crop yield utilising administered AI calculations, with the goal of providing a simple to use User Interface, increasing crop yield forecast precision, and investigating various climatic boundaries, such as overcast cover, precipitation, temperature, and so on. In the suggested framework, the state of MAHARASHTRA was chosen for implementation, and information was gathered using government websites such as www.data.gov.in. They used equations such as the Random Forest Algorithm to anticipate crop yields, and they established a website page to make it easy to use for everyone. The most well-liked

ET SHWETA: It has been deduced that the purpose of this research is to examine the many applications of AI in agriculture. Furthermore, using these processes, it is possible to select the proper crop, select land, and select the appropriate season. Naive Bayes and K-Nearest Neighbor are used in the calculations. Precision of execution is used in the calculations.

ET: AMIT KUMAR:

It is expected that this study will aid in anticipating crop arrangements, increasing production rates, and providing benefits to ranchers. Furthermore, using machine learning applications in farming to predict crop diseases, examine crop clones, and create different water system designs. Fake neuronal organizations were used in the calculations. The major challenge with brain organization is that determining which organization is optimal for the arrangement is challenging and requires trial. The second issue with brain organization is the dependency on technology. As the computation involves more calculations in reverse and forward, the amount of preparation required increases. Insight and time are required to ensure that the organization structure is acceptable. The suggested methodology also focuses on crop selection using both natural and financial variables. The framework also makes use of a monetary factor, which is the crop cost, which plays an important role if the yields are the same but the yield cost is different. Crop sequencing is another approach used by the framework, which provides a complete set of yields that can be developed throughout the season. Crop selection is also a focus of the proposed framework, which incorporates both ecological and financial factors. The framework also makes use of the monetary factor

ET: MANJULA It is expected that using rule-based mining, this research will aid in increasing crop yield rates. The paper employs affiliation rule mining to forecast agricultural yield. The k-Means Algorithm, bunching method, and derived affiliation rule mining were used in the calculations. The major stumbling block is that the article relies on affiliation rule digging to predict crop productivity. The problem with affiliation decision mining is that it occasionally generates an excessive amount of rules, reducing the precision of the expectation. Similarly, the concepts will vary significantly depending on the dataset, as will the results. The framework that has been proposed The majority of the discussion is around the topic of crop yield expectation, which plays an important role in crop selection since ranchers can select the crop with the highest yield. The framework uses affiliation rule mining to find the most profitable rules and crops. This framework focuses on the creation of an expectation model that may be used to anticipate agricultural yield in the future.

RAKESH KUMAR ET It is assumed that by using order procedures and looking at the boundaries, this paper will aid in enhancing agricultural yield rates.

The paper explains how to achieve the equivalent using various calculations. Bayesian computations, K-implies Algorithm, Clustering Algorithm, and Support Vector Machine are among the calculations presented. The stumbling block is that there may not be adequate precision and execution referenced in the study based on the stated computations. The paper is a research paper that just suggests using the calculations, but there is no proof provided in the publication. The crop decision technique used in this research is unusual in that it focuses on plants that may develop according to season. The suggested method settles crop (s) decisions based on anticipated yield costs and supported by boundaries (for example Environment, soil kind, water thickness, crop kind). It takes the crop, the planting time, the estate days, and the expected yield charge for the season as inputs and finds a succession of vegetation with the maximum day-to-day creation over the season.

3 IMPLEMENTATION STUDY

3.1DATA PRE-PROCESSING

By deleting the items that are transformed to integers, the raw data in the crop data is cleaned and metadata is appended to it. As a result, the data is simple to train. Take in all of the information. We first load the metadata into this, then attach the metadata to the data and replace the converted data with metadata in this pre-processing. The data will then be moved further and the unnecessary data in the list will be removed, and the data will be divided into the train and the test data To split the data into train and test, we'll need to import train test split from scikit-learn. This will assist the pre-processed data in splitting the data into train and test based on the weight specified in the code. The test and train are divided by 0.2 and 0.8, which equals 20% and 80%, respectively.

3.2 Model Creation:

We create datas into two models:

A) Training model

B) Testing model

The division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent respectively.

3.3 Model evaluation:

We apply the machine learning algorithm for testing part and get the accuracy of this model.

3.4 Prediction:

This module based on GUI part. we create a web page using bootstrap. The web page like (Nitrogen, Phosphorous, Potassium value, Humidity, Rainfall, Temperature). now we get the data from user to compare the dataset values. finally, it will predict for the Crop and soil to be planted.

3.5 Methodology:

Give the value of nitrogen, phosphorus, potassium, PH value, rainfall, humidity and temperature. We already trained the dataset. Our value compared to dataset and finally result will displayed what seed we cultivated that particular place.

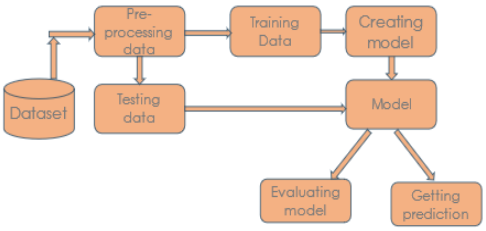

FIG1: - ARCHITECTURE Diagram

4 PROPOSED APPROACH

In the proposed system, we develop Prediction of the crop using the efficient algorithm. The challenge in it is to build the efficient model to predict the better crop Here in these projects we use machine learning algorithms like Voting classifier which is nothing but hybrid classification/ensemble of models. In our project the Voting classifier is an ensemble of models that are obtained from svm,randomforest and KNN. Which can enhance the accuracy and it can give a better prediction system.

4.1 ADVANTAGES OF PROPOSED SYSTEM

1)Predicting the better crop is the ultimate Aim of the project.

2)Early detection of problems and management of that problems can help the farmers for better crop yield.

3)For the better understanding of the crop yield, we need to study of the huge data with the help of machine learning algorithm so it will give the accurate prediction of crop and suggest the farmer for a better crop.

5 ALGORITHMS USED

5.1 Random Forest

Random forest is a supervised learning algorithm.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

Working of Random Forest Algorithm:
Step1 : First, start with the selection of random samples from a given dataset.
Step2 : Following that, this algorithm will create a decision tree for each sample. The forecast result from each decision tree will then be obtained.
Step 3: Each projected outcome will be voted on in this round.
Step 4: Finally, choose the prediction result with the most votes as the final forecast result.

5.2 KNN Model

- The KNN algorithm implies that items that are similar are close together. To put it another way, related items are close together.
- For example "Birds of a feather flock together."

It can be used to solve both classification and regression problems

5.3 Support Vector Machine

- The Support Vector Machine (SVM) algorithm is a simple yet effective Supervised Machine Learning approach that can be used to create both regression and classification models. The main goal of SVM is to construct the hyperplane and divide the So that datamining and categorization may be done successfully, divide the dataset into separate groups. SVM can be applied to real-world problems such as detecting tension in tweets.

5.4 VOTING CLASSIFIER
A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output. It simply aggregates the findings of each classifier passed into Voting Classifier and predicts the output class based on the highest majority of voting. The idea is instead of creating separate dedicated models and finding the accuracy for each them, we create a single model which trains by these models and predicts output based on their combined majority of voting for each output class.

Voting Classifier supports two types of votings.

1. Hard Voting: In hard voting, the predicted output class is a class with the highest majority of votes i.e the class which had the highest probability of being predicted by each of the classifiers. Suppose three classifiers predicted the *output class(A, A, B)*, so here the majority predicted *A* as output. Hence *A* will be the final prediction.
2. Soft Voting: In soft voting, the output class is the prediction based on the average of probability given to that class. Suppose given some input to three models, the prediction probability for class A = (0.30, 0.47, 0.53) and B = (0.20, 0.32, 0.40). So the average for class A is 0.4333 and B is 0.3067, the winner is clearly class A because it had the highest probability averaged by each classifier.

Now we combine all the model that are obtained from above three classifiers Random forest,KNN and SVM. In our project we used soft type voting that means here the output class is predicted based on the average of probability given to that class.

Convolutional neural network model:

- Fully linked networks are multilayer perceptron's in which each neuron in one layer is connected to all neurons in the next layer.
- CNNs are regularized versions of multilayer perceptron's.
- In CNN, there are three sorts of layers: input layer, hidden layer, and output layer.
- The hidden layer receives the input from the input layer.
- Depending on our model and data size, there could be a lot of hidden layers.
- The output of each layer is calculated by multiplying the output of the previous layer by the learnable weights of the layers in a matrix.

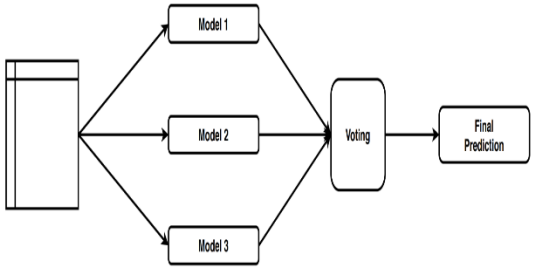

Fig 2:- Layered Convolution Neural Networks for crop disease prediction classification

The first layer, 1-CNN, is made up of 1-convculation. In this layer, the kernel size and 0-padding are used to create the layer and the temporal convolution process. The pooling operation is then executed when the relu activation functionality is applied to the layer. Global maximum pooling is used here to reduce the dimensionality of the data. The pooling output is used as an input for the fully connected convolution layer that creates the single value. This is used to perform the sigmoid activation function.

When a photograph of an insect is submitted, we use Convolutional neural networks to identify the type of bug in our project. Because we can only anticipate the insecticides for that insect if we know what kind of insect it was in the first place. In our project, we used three convolutional layers.

The image is 64x64 pixels in size.

A convolutional neural network is a feed-forward neural network that processes data in a grid-like structure to evaluate visual pictures. A ConvNet is another name for it. To detect and classify items in an image, a convolutional neural network is employed. Multiple hidden layers in a convolution neural network aid in the extraction of information from an image.

CNN's four most significant tiers are:

1.Convolution layer

1. ReLU layer
2. Pooling layer
3. Fully connected layer

Convolution Layer

This is the initial stage in obtaining useful information from a photograph. The convolution action is performed by many filters in a convolution layer. Every image is viewed as a pixel value matrix.

ReLU layer

The rectified linear unit is abbreviated as ReLU. After the feature maps have been removed, they must be moved to a ReLU layer.

ReLU performs an element-by-element procedure, setting all negative pixels to zero. It causes the network to become non-linear, and the result is a rectified feature map.

Pooling Layer

Pooling is a down sampling process that decreases the feature map's dimensionality. To create a pooled feature map, the rectified feature map is now sent via a pooling layer.

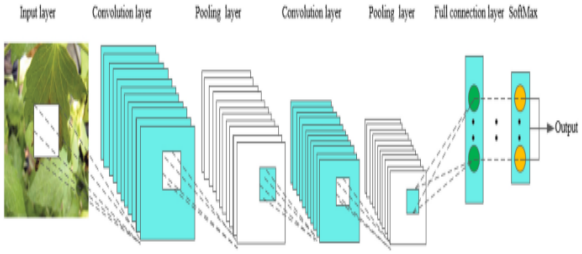

Fig 3:- Proposed CNN model for Disease prediction

6 RESULTS AND EVOLUTION METRICS

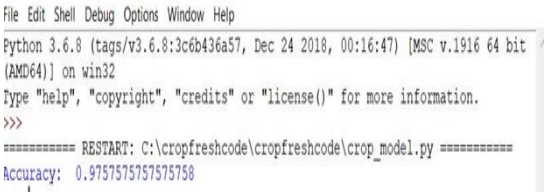

Fig 4: - voting classifier accuracy which was achieved on taring and testing dataset was 97.57

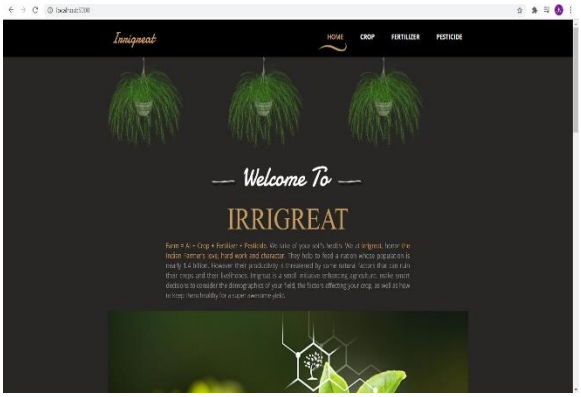

Fig 5 home page for prediction system

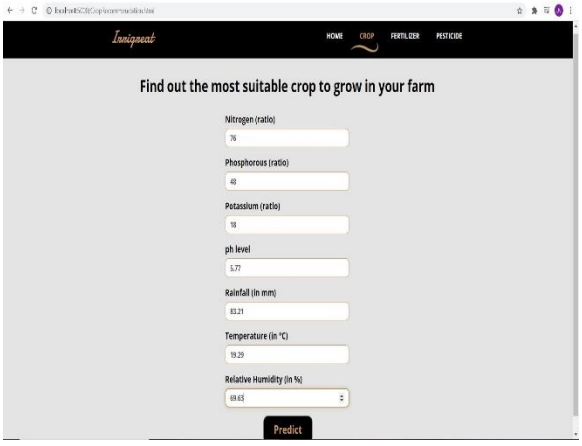

Fig6: -input form for prediction for crop using parameters of crop yield such as nitrogen, ph, potassium etc.

Fig7: -predicted crop based on inputs here the recommendation system was implemented

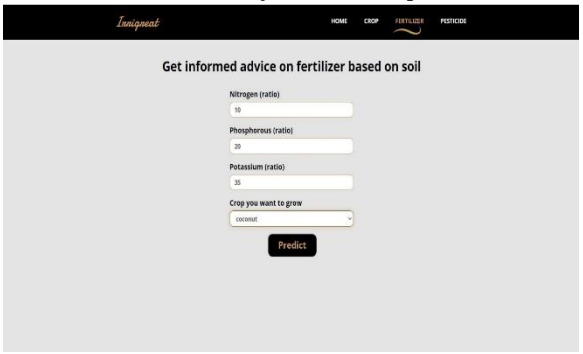

Fig8: -input form prediction for fertilizers

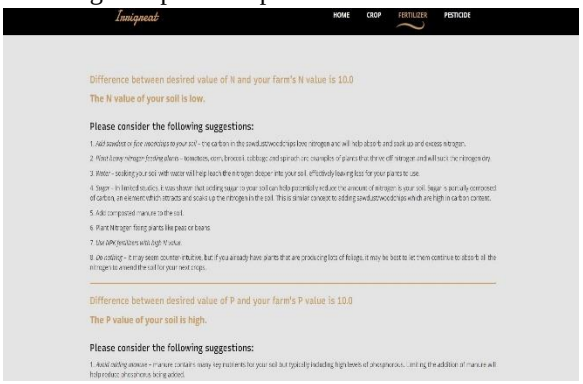

Fig9:- recommendation system based on the fetrizers

Fig10: - input image for recommendation for pesticides based on pest

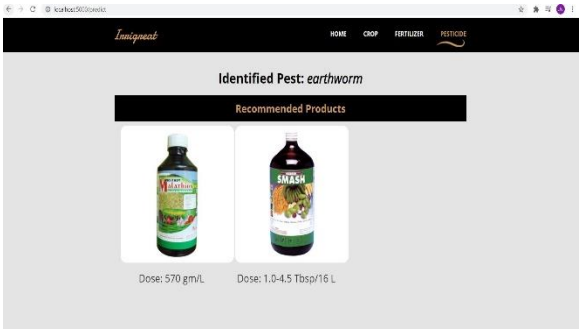

Fig10: - prediction of types of pest as well recommendation for prevention of disease and doses to be used for the pest

CONCLUSION

Voting classifier, which is nothing more than an ensemble of models, is used in the proposed work to create a crop prediction framework. The models developed from svm, random forest, and knn are used in our project voting classifier ensembles. Our project improves the accuracy of crop forecasting. As a result, the framework will assist farmers in reducing the obstacles they face and preventing them from committing suicide. It will serve as a conduit for providing farmers with the information they need to earn a high return and, as a result, increase benefits, lowering self-destruction rates and reducing his obstacles.

REFERENCES

[1] Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara, Mansing Rathod, “Crop yield prediction using machine learning” International Journal of Science and Research ,April 2020.

[2] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate, “Crop Prediction System using Machine Learning Algorithms”, International Research Journal of Engineering and Technology, Feb 2020.

[3] Ramesh Medar, Shweta, Vijay S. Rajpurohit, “Crop Yield Prediction using Machine Learning Techniques”, 5th International Conference for Convergence in Technology, 2019.

[4] Trupti Bhange, Swati Shekapure, Komal Pawar, Harshada Choudhari, “Survey Paper on Prediction of Crop yield and Suitable Crop”, International Journal of Innovative Research in Science, Engineering and Technology, May 2019.

[5] E. Manjula, S. Djodiltachoumy, "A Modal for Prediction of Crop Yield "International Journal of Computational Intelligence and Informatics, March 2017.

[6] Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, "Crop Selection Method Based on Various Environmental Factors Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Feb 2017.

[7] Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India., May 2015.

[8] Rajshekhar Borate., "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Software Engineering, Vol. 3, Issue 1, pp: (34-37), April 2016.

[9] D Ramesh, B Vishnu Vardhan, "Analysis of Crop Yield Prediction using Data Mining Techniques", International Journal of Research in Engineering and Technology (IJRET),Vol.4, 2015.

[10] S.Veenadhari, Dr Bharat Misra, Dr CD Singh.2019."Machine learning approach for forecasting crop yield based on climatic parameters.".978-1-4799-2352- 6/14/$31.00 ©2014 IEEE.

[11] Igor Oliveira, Renato L. F. Cunha, Bruno Silva, Marco A. S. Netto.2018."A Scalable Machine Learning System for PreSeason Agriculture Yield Forecast.".978-1-5386-9156- 4/18/$31.00 ©2018.

[12] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein."Prediction of Crop Cultivation".978-1-7281-0554-3/19/$31. 00© 2019 IEEE.

[13] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty.2017."Agricultural Production Output Prediction Using Supervised Machine Learning Techniques".978-1-5386-3831-6/17/$31.00 ©2017 IEEE.

[14] G Srivatsa Sharma, Shah Nawaz Mandal, Shruti Kulkarni, Monica R Mundada, Meeradevi.2018."Predictive Analysis to Improve Crop Yield Using a Neural Network Model".978-1-5386-5314-2/18/$31.00 ©2018 IEEE.

[15] Rashmi Priya, Dharavath Ramesh.2018."Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model". 978-1-5386-5314- 2/18/$31.00 ©2018 IEEE.