Crop and Fertilizer Prediction and Disease Detection Using Data Science

M GOUTHAM[1], KONAGANTI PRAVALYA[2], G.NAGA VAMSI[3], ROYYURU SRIKANTH[4]
[1, 2, 3, 4] *Dept. Of Computer Science and Engineering, Vardhaman College of Engineering, Telangana, India*

Abstract*—One of the key industries that affect a nation's economic development is agriculture. The bulk of people in countries like India relies on agriculture for their livelihood. For growth and increased output, plants require nutrients. Plants can get the nutrients they need from fertilizers, manure, and soil. By putting the plant under both abiotic and biotic stress, climate change and global warming were found to be the primary causes of crop loss. One of the main areas of focus for academics globally is figuring out the causes of and solutions to the issues associated with crop loss [1], it is not only causing crop loss but is now affecting food production and crop prediction, which will have a detrimental effect on farmers' economies by lowering yields and making them less skilled at predicting future harvests. This research project helps inexperienced farmers (Tech farmers) plant the right crops by utilizing Data Science, one of the most advanced technologies in mining and forecasting. To produce more agricultural products with less waste, the agricultural sector needs technological advancements. Therefore, our major goal is to make it simple for farmers or other users to work on their farms by developing a website that includes crop and fertilizer forecasting as well as plant disease detection. Overall, the integration of crop prediction, fertilizer prediction, and disease detection using Data science can improve agricultural sustainability, productivity, and profitability. By leveraging these technologies, farmers can reduce crop loss due to disease outbreaks, increase crop yield and quality, and minimize the negative impact of agriculture on the environment.

Indexed Terms- Decision Trees, Flask, Forecasting, Random Forests, Residual Network, Support Vector Machine, Web Technologies.

I. INTRODUCTION

Agriculture is the backbone of India's economy, and with over 60% of the country's population engaged in farming, it is critical to ensure that the sector is sustainable and productive. Unfortunately, many farmers lack access to the knowledge and resources necessary to optimize their yields and minimize the environmental impact of their practices.

We have developed a comprehensive system that leverages Data science to address the challenges faced by farmers in India. Farmers often lack sufficient knowledge about soil nutrients and crop management, which can result in soil depletion, lower yields, and crop diseases. Our system provides personalized recommendations for crop selection and fertilizer application, considering factors such as soil type, climate, and crop history. This allows farmers to optimize their yields while minimizing environmental impact.

Due to plant diseases, plant yields are significantly reduced, which results in significant economic losses. [1]. To overcome this our system includes plant disease prediction models that can accurately identify potential outbreaks and provide timely recommendations for disease prevention and treatment. This is especially important for smallholder farmers who depend on healthy crops for their livelihoods and may not have access to traditional agricultural extension groups or local plant clinics.

By making this information available online, we can help farmers overcome the barriers to accessing resources and improve food security for the country.

II. LITERATURE SURVEY

To detect undernourishment in the case of crops, a variety of methods have been employed. The detection of diseases, the use of fertilizer, and crop forecasting have all been the subject of several papers.

Ashwani Kumar Kushwaha - Outlines fertilizer and crop yield prediction techniques and recommends an appropriate crop to boost farmer profit and the standard of the agricultural industry. To get a large amount of data (soil and meteorological data) for crop production forecasts, this article employs the Hadoop platform and an agricultural algorithm. The repository's data may be used to anticipate which crops will be best in a certain circumstance and to boost crop quality[2].

Anand H. Kulkarni - Created a system with a detection rate of up to 91 per cent that uses a range of image processing techniques, for example, the usage of a Gabor filter for extraction of features and an ANN-based learner for classification [3].

E. Manjula - Created a model using Naive Bayes, Decision Trees, and a blend of Naive Bayes and Decision trees to look into the soil's nitrogen, phosphorus, potassium, calcium, magnesium, sulphur, iron, and zinc content. Based on the accuracy and execution time, the classification algorithms' performance is contrasted [4].

V. G. R. R. Devi, K. Neha, and Vaishnavi. - With this suggested approach, crop production may be predicted using historical data that takes into account elements like temperature, humidity, ph, rainfall, and crop name. The greatest variety of crop varieties would be covered throughout all Indian districts under this approach. Using the recommended method, we can predict the best crop depending on the weather in the field. Crops may be predicted using the decision tree and random forest algorithms. To find the most precise value, the random forest method was used. The benefit from agricultural production grew with more precise discoveries. [5].

Bhagyashri Dadore, D. Anantha, and Aarti Watekar. This system's design was based on three variables: soil types, soil features, and crop production statistics. Based on these variables, the system suggested to the farmer which crop would be best to grow. The machine learning techniques used by this suggested system included the random forest, CHAID, K-Nearest Neighbor, and Naive Bayes. We can forecast certain crops under specific meteorological circumstances, as well as state and district values, by using the suggested approach. [6].
Avinash Devare, Mitalee Pendke, Pooja Shinde, Rohit Kumar, Rajak, and Ankit Pawar - Based on the soil database, this suggested approach is utilised to identify certain crops. This suggested system was successful with a wide range of crops, including groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, and coriander, as well as with a wide range of soil properties, including depth, texture, pH, soil colour, permeability, drainage, water holding, and erosion. This system's recommendation of a crop for site-specific factors was made with accuracy and efficiency the support vector machines (SVM) classifier, the ANN classifier, the random forest classifier, and the naive bayes classification algorithm are just a few examples of the machine learning classifiers that were used. Farmers would benefit from this research's assistance in increasing agricultural output, preventing soil erosion on cultivated land, using water resources wisely and decreasing the usage of pesticides in agricultural cultivation[7].

III. PROPOSED APPROACH

The purpose of this project is to reduce the time and effort required by farmers by developing a website that recommends crops and fertilisers based on various factors, including rainfall, ph, state, district, nitrogen, phosphorus, and potassium. There have been several proposals to use deep learning techniques to detect leaf diseases as It is particularly difficult to detect illnesses in the agriculture sector. So, we trained the model to forecast the potential plant illness using CNN's RESNET.

IV. REQUIREMENTS

I. User Requirements
I.I For Crop Prediction

- Nitrogen value (N)

- Potassium value(P)
- Phosphorous value(K)
- Rainfall in mm
- PH value
- Humidity
- State and District

I.II For Fertilizer Prediction

- Nitrogen value (N)
- Potassium value(P)
- Phosphorous value(K)
- Crop name

I.III For Plant Disease detection

- Image of the leaf.

II. REQUIREMENTS FOR SOFTWARE AND HARDWARE

- Windows 7 and later operating systems.
- Python 2.7 and later is the required programming language.
- Anaconda Jupyter Notebook is the platform.
- Tensorflow, OpenCV, PIL, tkinter, OS, and SKlearn are supporting libraries.
- A processor running at 2.5 GHz or above.
- 4 GB more than Memory is recommended.

IV. ALGORITHMS AND FLOWCHARTS

- SVM, or supervised machine learning, is a technique that may be used to solve classification and regression issues.
- When there are multiple explanatory factors, logistic regression is used to calculate the odds ratio.
- In a decision tree, each internal node represents a "test" on such a characteristic, and each branch represents the test's outcome, so each leaf node represents a class label. The structure mimics a flowchart.
- Random forest is a well-liked supervised machine learning method for addressing classification and regression problems. It constructs decision trees from the range of samples that use the averages for regression as well as the simple majority for classification.
- The computing capacity of boosted tree algorithms is pushed to the limit by the XGBoost approach for gradient boosting. It is very accurate and scalable. Its primary goal was to speed up computation and enhance the effectiveness of models used for machine learning.
- ResNet is a neural network convolution (CNN) design that overcomes the "vanishing gradient" issue to produce networks with dozens of convolution layers that outperform shallower networks.

Chart -1: Flowchart for crop and fertiliser recommendations.

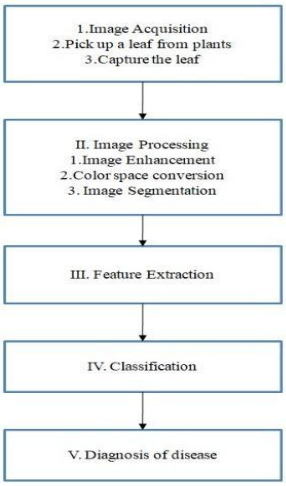

Chart -2: Flowchart of Plant Disease Detection

V. EXPERIMENTAL SETUP DESIGN

In this project we are trying to combine technology and agriculture to help the farmers in producing good crops. So, we are building a website which has three models they are crop prediction, fertilizer prediction and plant disease detection. In crop and fertilizer prediction we used a dataset which has 2200

instances and Eight attributes crop, pH, humidity, phosphorus, potassium, rainfall, and nitrogen This data is cleaned and using Sklearn, we divide it into training and testing sets that are 80/20 each. We then develop a prediction model using a variety of methods, including SVM, XGBoost, logistic regression, decision trees, and random forests. The other model is plant disease detection in this we used a large data set which has 10 different pants such as potato, tomato, strawberry, grape, pepper, orange, and many others we used the torchvison, a library of python to convert the image pixels into 0-1 range from 0-255 so we can give it to neural networks, we used trochsensors for storing the values since torch sensor can be on GPU and GPU is best for image data. We divided the images into batches and gave them to the model and before dividing them into batches the images were shuffled. This helps in better training. For training, we used the ResNet architecture of CNN. This is the deepest architecture in CNN and it also solves the vanishing gradient problem and which improves the accuracy of the model.

MODEL DESIGN

The models we developed are deployed using pickle in python and these pickle files are used in the website. For designing the user interface of the project we used html, css, WordPress, and javascript. The website is deployed using python flask, it is run in the local host of the system.

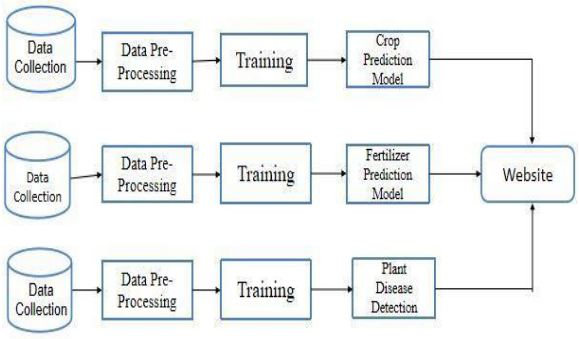

Chart -3: Proposed Model Design

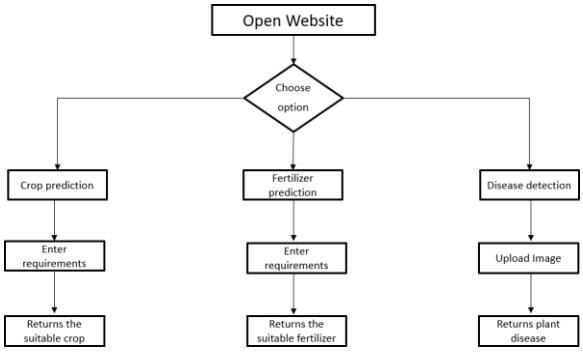

Chart -4: Proposed Model Flow Chart

IMPLEMENTATION

In crop and fertilizer prediction we use a dataset which contains 2200 instances and N, P, K, Ph, Rainfall, Humidity, Temperature, and Crop Name make up the first eight properties. In the first step of pre-processing, we changed the name of the crops and then we changed all the words to lowercase and removed all the redundancies. The data was then divided into 80-20 per cent of the overall training and 20% of testing sets to use the sklearn training test split function. After that, we trained the data using the XGBoost method, Decision Tree, Gaussian Naive Bayes (GNB), Support Vector Machine (SVM), Logistic Regression, Random Forest, and SVM algorithms. We used the aforementioned techniques to produce a prediction model. As the large data era develops, neural networks based on convolution (CNNs) with far more hidden layers have a more intricate network structure and stronger feature learning and feature expression capabilities than traditional machine learning techniques. Since its introduction, the convolution neural network framework taught by the algorithm for deep learning has shown excellent results in several huge identification tasks in the area of computer vision. The origins of deep learning and convolutional neural networks, the fundamental system models, convolution extraction of features, and convolutional neural networks (CNN network pooling operations are covered in the first section of this article. After that, it is described the current state of research and the direction of development of a convolutional neural network model for image classification based on deep learning, which is primarily introduced from the standpoint of typical network properties.

VI. RESULTS AND DISCUSSION

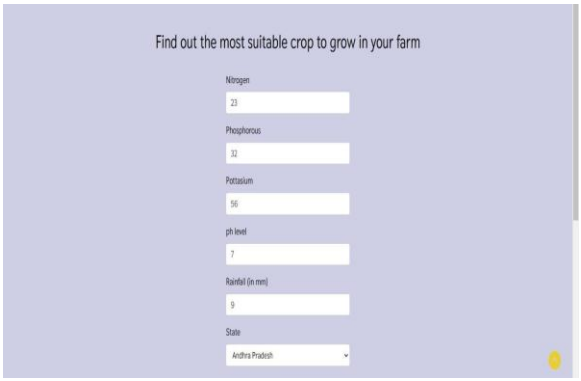

Fig -1: Inputs for crop prediction

Fig -2: Output for crop prediction

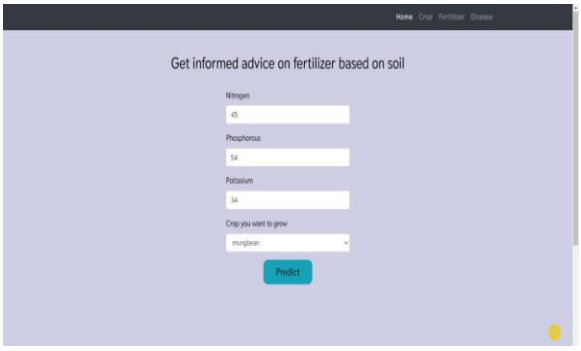

Fig -3: Input for fertilizer prediction

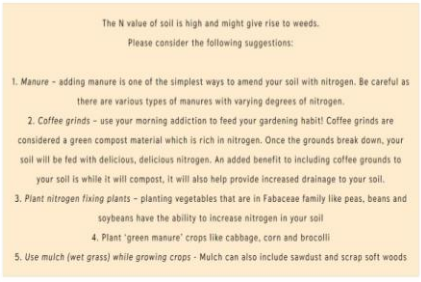

Fig -4: Output of fertilizer prediction

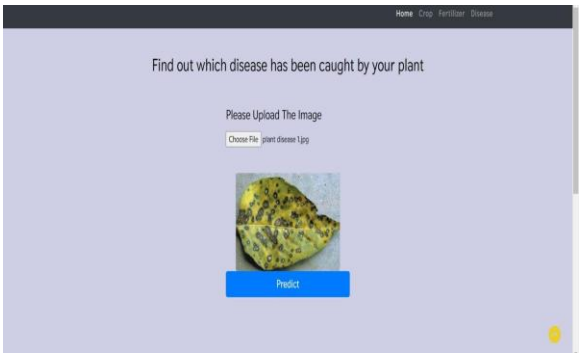

Fig -5: Input for disease prediction

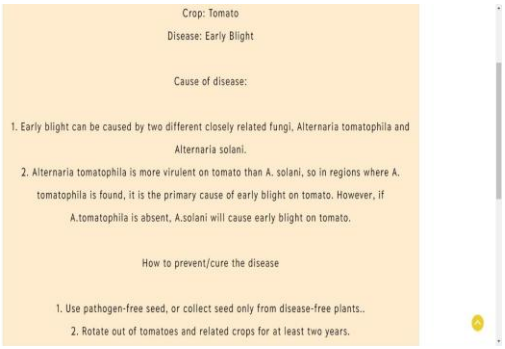

Fig -6: Output of disease prediction

Table -1: Crop and Fertilizer prediction

S.no.	Methodology	Performance	Accuracy
1	Decision Tree	93.6%	92.9%
2	GNB	99.5%	98.8%
3	SVM	98.8%	98.4%
4	Logistic Regression	96.5%	95.5%
5	Random Forest	99.5%	98.8%
6	XGBoost	99.3%	99.1%

Table -2: Plant disease detection

S.no.	Methodology	Performance	Accuracy
1	CNN	98.2%	97.5%

CONCLUSION

Our work aims to combine the benefits of technology with agriculture. Our website predicts the crops to be grown and fertilizers to be used. The website can also be used to predict the disease of the plant and suggest how to cure the disease. In this project, we are not physically collecting the input requirements, but rather depending on the farmer or the user for the data. In future, we can create an IoT-based product which can be kept on the farms for automatically collecting the required data and updating it to the database.

REFERENCES

[1] S. Roy, R. Ray, S. R. Dash, and M. K. Giri, 2021. Plant disease identification with a focus on dimensionality reduction using machine learning technologies. A Machine Learning Perspective on Data Analytics in Bioinformatics, pp. 109–144.

[2] R. Balodi, S. U. N. I. N. A. Bisht, A. Ghatak, and K. H. Rao, 2017. Diagnostic problems and technical improvements for plant diseases. Pages. 275-281 in Indian Phytopathology, 70(3).

[3] Agro algorithm in Hadoop for crop production prediction, Kushwaha, A.K. and Bhattacharya, S., 2015. IJCSITS, 5(2), pp. 271-274. International Journal of Computer Science and Information Technology & Security.

[4] Detecting plant diseases using image processing approaches, Kulkarni, A.H. and Patil, 2012. pp. 3661–3664 in International Journal of Modern Engineering Research, 2(5).

[5] K. Neha, V. Spandana, V. S. Vaishnavi, Y. Jeevan Nagendra, and V. G. R. R. Devi. supervised machine learning approach for predicting crop yield in the agricultural sector Pages 736-741 of the 2020 Fifth International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020.

[6] Bhagyashri Dadore, D. Anantha, Reddy, and Aarti Watekar. "A machine learning-based crop recommendation system to enhance agricultural production in rate area." 6(1) (2019), 485-489, in International Journal of Scientific Research in Science and Technology.

[7] Avinash Devare, Mitalee Pendke, Rohit Kumar, Ankit Pawar, Pooja Shinde, and Rajak. "A crop recommendation system employing machine learning to enhance agricultural productivity." 4, no. 12 (2017): 950-953 for the International Research Journal of Engineering and Technology.