International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
Article Received: 27August 2022 Revised: 05October 2022 Accepted: 20October 2022

CROPUP – A Crop Yield Prediction and Recommendation System with Geographical Data using DNN and XGBoost

Sobhana M[1], Smitha Chowdary Ch[2], D.N.V.S.L.S. Indira[3], Konduru Kranthi Kumar[4]
[1]Senior Assistant Professor
Depatrment of Computer Science and Engineering,V R Siddhartha Engineering College,Vijayawada, Andhra Pradesh, India.
sobhana@vrsiddhartha.ac.in
ORCID:0000-0001-5938-5740
[2]Associate Professor,
Department of Computer Science and Engineering Data Science, Malla Reddy Engineering College for Women, Telangana,A.P, India..
smithacsc@gmail.com
*ORCID:*0000-0002-9775-9632
[3]Associate Professor
Department of Information Technology,Seshadri RaoGudlavalleru Engineering College,Gudlavalleru, Andhra Pradesh, India.
indiragamini@gmail.com
*ORCID:*0000-0003-1631-1156
[4]Assistant Professor
Department of Information Technology,Vasireddy Venkatadri Institute of Technology ,Namburu, Pedakakani, Guntur, Andhra Pradesh, India.
kk97976@gmail.com
ORCID:0000-0002-8419-7386

Abstract—Agricultural management is significant in a populous country like India. Farmers must have advance knowledge about predicted crop production and crop condition within particular area to make economic and farming decisions. To generate yield, we consider factors like temperature, humidity, pressure, NDVI values, Latitude, Longitude etc. When cultivating a particular crop on a specific type of soil, there are a number of factors to be considered. A crop recommender system considers soil properties such as N, P, and K, as well as other factors like rainfall, humidity, and pH levels, to choose the best crop for the farm. This paper presents a predictive algorithm that would estimate crop yield using deep neural networks with geographical data. A recommendation system was built using machine learning algorithm like Xgboost to recommend the suitable crop. A user interface named CROPUP has been developed to scale up crop productivity and efficiency using the proposed algorithms.

Keywords-Crop Yield, Deep Neural Networks, Geographical data, Recommendation, Xgboost

I. INTRODUCTION

The agricultural market is growing as the global population increases. The annual cycles of agricultural phenology and the reliance of crop production on climate, soil, and weather conditions make it challenging to regulate farming activities. A large majority of agriculturalists do not make the desired profit as a result of these issues. Farmers need immediate guidance on crop Eco perspectives. An analysis is required to assist farmers in boosting their profitability. Yield estimation is one of agriculture's most significant challenges. Crop yield is something every agriculturalist wants to understand more about. When making agricultural suggestions, it is vital to have a thorough understanding of crop production history. Crop yield prediction used to be done by statistically evaluating the cultivator's prior knowledge of the crop. Furthermore, there is a lot of farming data, and manually analyzing it is time-consuming. Producers often used forecast output by looking at the previous year's crops. Crop prediction with Machine Learning employs a variety of approaches or algorithms, and these algorithms can predict crop yields [1]. In application areas, machine learning is now being utilized to increase crop yield prediction accuracy. Machine learning has come a long way, but its utility in data-driven applications is still limited. Its precision is influenced by the model's representativeness and information quality, as well as the input elements in the collected data. As a result, deep-learning approaches have been used to forecast crop yields using variables such as latitude, longitude, temperature, pressure, humidity, apparent temperature, and normalized difference vegetation index

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

(NDVI). The sequential model for crop yield estimation is based on real data and does not include any attributes that were created manually [2].

To generate a precise yield forecast from known weather conditions, a sequential model was constructed to train nonlinear and complicated connections between input factors using geographical data. This study proposes the sequential technique for simulating climatic changes and climate factors including temperature, pressure, humidity, and NDVI. The sequential model is one of the methods in Deep Neural Networks. DNN has one input layer, several hidden layers, and one output layer. The result of one layer is transmitted as an input to the next layer, and so on until the output layer is reached [3]. Every layer is made up of many neurons that process the input and generate output. It enhances agricultural yield prediction accuracy by including a variety of climate-related variables.

Farmers can connect to the planned crop recommendation system via a web interface. Algorithms based on Machine-learning allow for the selection of the most lucrative crop. Machine learning methods such as Random Forest, Support Vector Machine, Naive Bayes, Decision Tree, and the famous ensemble algorithm Xgboost are used to make recommendations. N, P, and K values, as well as other elements like rainfall and humidity, are used to predict the crop. A decision tree uses attributes to classify data. Decision nodes and decision leaves make up a tree. Nodes can contain two or more branches, each of which indicates a value for the characteristics being tested, and the nodes of the leaf generate a homogeneous result. Random forest is an ensemble algorithm for regression, classification, and for problems that generate a set of call trees at any given time and outputs the category. The practice of call trees overfitting to its coaching set is corrected by random call forests. XGBoost is a technique of ensemble learning. Gradient Boosted decision trees are implemented in XGBoost. The XGBoost classifier produces the best results, with a 99.31 % accuracy; hence it was used to create the final model. Farmers will be able to choose the ideal crop for their land with the help of this method.

II. Related work

Kale [4] used neural network regression methods to look at the productivity of various crops. Crop data obtained are used for the whole year including seasons like summer, kharif and rabi. Only data from Maharashtra have been included in the dataset. Utilizing the RMS prop optimizer, the output was initially 45% accurate, but the model was later enhanced to 90 % by increasing the number of layers, altering the weight, and bias, and using the Adam optimizer

ANN model was utilized by [5]. An ANN is used to estimate agricultural yields by sensing several climate and soil parameters. It considers soil type, rainfall, pressure, water depth, temperature, humidity, nitrogen, potassium, phosphate, and organic carbon. In this work, the impact of these aspects is researched and analyzed. Crop production rate has been connected to meteorological conditions, soil composition, and soil type. Based on a predicted crop yield rate, this study also recommends an appropriate crop. Crop production rates are being modeled and predicted using artificial neural networks, which improve crop yield forecast accuracy.

Jianqiang [6] used acquired LAI from aerial photography to estimate yield using wheat crop. To reduce yield forecast error and hence enhance data efficiency, the

Savitzky-Golayphiltre was applied to optimize the NDVI pattern. The Gaussian model was used to approximate typical crop LAI at each growth point in this study to obtain an average LAI at each growth point. There were several causes of mistakes in the crop yield estimation technique, including inheritance and aggregation problems. Data preparation mistakes were one of the key issues of the production regression coefficients.

The purpose of Chuan [7] was to see if drought-related yield reduction and protection were particularly important in precision agriculture. Soil humidity was among the key factors of crop yield in dry and semiarid conditions. Despite its time and geographical resolution limitations, this project is only the first step in estimating crop productivity using temperature vegetation dryness index (TVDI). Furthermore, in many real-world situations involving fluctuating soil moisture and structure, the dry edge inside the pixel distribution picture of land surface temperature (LST)-NDVI is not linear. Inaccuracies in the prediction model may result from these variables. Increases in the aforementioned variables were expected to improve accuracy.

Mythili[8] concentrated their efforts on detecting flood damage in agricultural fields. These losses are directly proportional to the influence it has on the crops, necessitating the proper prediction of data to obtain more precise outcomes. The researchers determined flood damage to crops using the Moderate Resolution Imaging Spectroradiometer (MODIS) weekly NDVI product. For data analysis, the time-series idea was utilized to ensure that the weather change concerning time intervals was largely the same.

Pavani [9] provided data from numerous Telangana counties from the TS Development Society. Crop production was anticipated using the Nearest k technique. For a particular time, several weather conditions and soil requirements were considered nearest neighbors. In conclusion, K-nearest neighbor (KNN) was found to produce more exact and productive harvest yields.

According to Mahagaonkar[10], crop yield prediction can be accomplished by implementing crop yield data, nutrients,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

and location data. The Random Forest and Support Vector Machine (SVM) classification techniques are used to process these data. For fertilizer recommendations, appropriate crops and required fertilizer for each crop are suggested. It also shows weather information for a given location, which is provided by third-party Application Programming Interface (API). SVM is good at predicting crop yields, with a 99.47 % accuracy.

Rakesh [11] introduced a technique called Crop Selection Method (CSM) for selecting the crop planting sequence across a season. Random Forest, SVM, and KNNs are among the machine -learning approaches researched and contrasted. The proposed technique, which is based on yield rate prediction driven by weather, soil type, water density, and crop type, resolves crop selection. The main advantage of this model is that it maximizes crop net production rate across the season, resulting in the maximum economic growth.

The research of Shanmuganathan[12] has aided in making a connection between temperature variations and a vineyard in New Zealand (NZ). The attributes with similar values were effectively grouped with the support of Neural Networks A Chi-square test was used to analyze the impact of the system factors. Temperature fluctuations, in addition to other factors such as humidity, wind speed, and precipitation, were found to play a significant role in deciding the production and quality of the crop.

Dr. Suresh [13] advised using an information mining strategy to aid ranchers in determining the soil quality by conducting tests on its various boundaries and recommending crops based on the results. To boost the effectiveness of the Harvest Suggestion Framework, the framework makes use of the Help Vector Machine's arrangement calculation. The framework plots the soil and yield data to forecast a list of feasible harvests for the soil, as well as provides information on supplements that are deficient in the soil for the specific harvest.

Lavreniuk[14] employed CNNs to classify images and created a crop yield prediction model utilizing RGB and NDVI data from UAVs and presented a multi-level DL based on RNNs and CNNs. To simulate their yield outputs, the researchers extracted geographical and temporal variables from inputs such as sensed time-series data and soil properties. The study's experimental findings demonstrated its superiority over alternative approaches. The model achieves accuracy of 85% for crops like maize and soybeans.

Oikonomidis[15] conducted a literature survey for identifying and assessing the most appropriate studies. The Convolutional Neural Network (CNN) was found to be the most widely used algorithm and to have the best Root Mean Square Error (RMSE) which helps in predicting data accurately .The researcher found 456 supporting documents after utilizing selection and quality-assessment measures to the appropriate literature, from which 44 primary publications were identified for additional investigation. The primary investigations were utilized to investigate and synthesize the key motives, target crops, methodologies used, features utilized, and data sources used.

An intelligent crop recommendation system was introduced by Doshi [16] based on environmental data, farm geolocation, and soil properties, the suggested system recommends a crop according to particular season . Another method for rainfall prediction using Linear Regression was also created, which predicts rainfall for the following 12 months. Crop recommendation techniques comprised Neural Network, Decision Tree, Random Forest, and K-nearest neighbors. Their proposed approach includes a map - visualization function that aids farmers in deciding which crop to plant.

III. Methodology

The proposed method predicts the yield of a crop based on both climatic and geographical data. The sequential model which is a Deep Learning Model is created and trained with the above data to forecast the yield. A crop recommender system is also designed to suggest a suitable crop for the soil based on the soil nutrients using different machine - learning techniques to maximize the crop yield.

A. Dataset Collection

Data were obtained from a variety of sources during the data collection phase, and datasets were created. The climate data and geographical data were gathered from different government websites. There are several Internet summaries accessible, including aps.dac.gov.in, imd.gov.in, data.gov.in, aerialintel.blob, etc. The sample dataset used for yield prediction is shown in Table 1. The crop recommender dataset consists of nitrogen, phosphorous, potassium, temperature, humidity, pH, and rainfall values which are collected from kaggle.com (the sample dataset is shown in Table 2). Parameters considered for crop recommendation are represented in Figure 1.

a) GeographicalData

NASA and other space organisations provide data about the earth's surface. The majority of vegetation indices are data-condensed, which enables them to precisely and economically characterize natural vegetation cover and development conditions. These vegetation indices (Vis) are based on NASA's Terra/MODIS products, including NDVI, EVI.NDVI is indeed a good metric, according to extensive research. The NDVI has proven to be one of the most effective measures of agricultural growth conditions. It is a combination of red (Rr) and near-infrared (Rn) reflectance measurements.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

TABLE I. SAMPLE DATA SET OF YEILD

Latitude	Longitude	Apparent Temp min	Apparent Temp max	Humidity	Pressure	Temp min	Temp max	NDVI	Yield
93.599	7.525	20.67	31.46	0.91	0.0027	23.38	31.27	0.55935	357.8
92.665	11.528	20.25	30.71	0,94	0.0012	23.4	30.92	0.31392	317.4
79.72	14.435	20.4	31.02	0.75	0.0027	22.97	30.33	0.21197	144.5
95.943	28.955	20.15	30.68	0.67	0.0014	23.06	31.1	0.78458	465.9
81.392	16.887	20.33	30.73	0.69	0.0016	23.47	30.84	0.8159	256.8
83.979	18.571	20.41	30.7	0.74	0.0012	23.26	31.14	0.32868	179.4
78.774	14.451	21	31.06	0.69	0.0026	23.31	30.52	0.56166	277.2
79.004	13.457	19.97	30	0.94	0.0027	23.43	31.33	0.78730	322.1
96.345	27.37	19.72	30.06	0.67	0.0027	23.29	30.24	0.78792	465.8
80.793	16.291	20.76	31.11	0.91	0.0013	23.45	30.86	0.77231	229.4
79.51	15.611	20.35	31.14	0.73	0.0019	23.15	30.48	0.35005	196.1

b) Climate Data

The Indian Meteorological Department (IMD) provides daily, monthly, and annual reviews of climate factors in India, including minimum temperature, maximum temperature, apparent temperature, dew point, precipitation, latitude, longitude, pressure, wind speed, visibility, cloud cover, etc. Kaggle also contains different datasets regarding climate and soil.

TABLE II. SAMPLE DATASET OF CROP RECOMMENDATION

N	P	K	Temperature	Humidity	Ph	Rainfall	Label
90	42	43	20.87974	82.00274	6.502985	202.9355	Rice
85	58	41	21.77046	80.31964	7.038096	226.6555	Rice
68	41	16	21.77689	57.80841	6.158831	1.2.0862	Maize
93	41	17	25.62172	66.50415	6.047907	105.4655	Maize
84	55	38	26.87484	79.78725	6.956683	173.1017	Jute
80	45	42	23.14265	74.9974	7.380396	151.9035	Jute
89	28	33	26.44414	53.83876	6.993236	175.3723	Coffee
112	39	29	26.12492	63.37479	6.726529	147.8035	Coffee

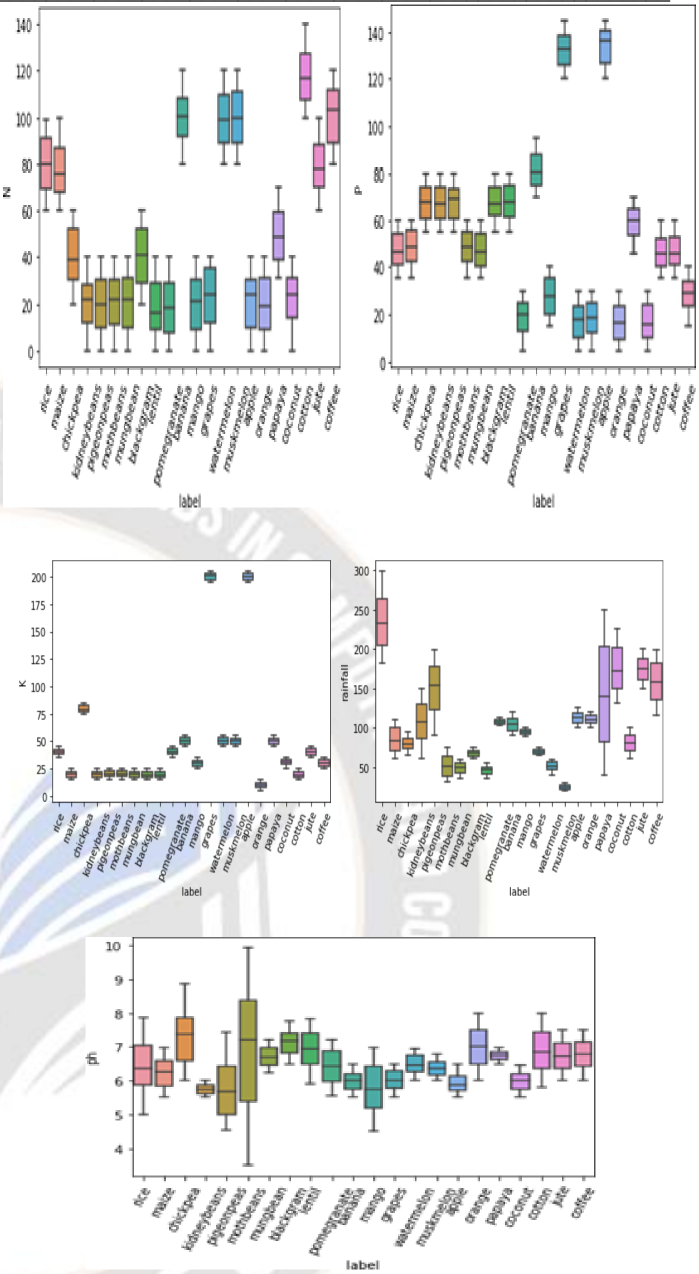

Figure 1. Box-plots of features considered for crop recommendation

B. Data Cleaning

When combining various data sources, there are various chances for data to be duplicated or wrongly categorized. If the data is incomplete, incorrect, it can disrupt operational efficiency. Data cleaning is the process of removing incorrect, corrupted, incorrectly structured, redundant, or insufficient information from a dataset. The data are preprocessed before being used. It refers to all of the modifications that the raw data undergo before being fed into the deep - learning model. Preprocessing is also necessary to expedite training. Typically, the received data contain a large quantity of noise from a variety of sources. Data preprocessing is indeed a method for transforming clumsy data into data clean. Data smoothing is a

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

data pre-processing strategy that removes noise from a data set . The data can be smoothed using the binning method. The data must be efficiently arranged to acquire better outcomes from the employed model in machine-learning algorithms.

a) *Feature Selection*

Feature selection is used to identify the significant features before training the model, as represented in Figure 2. It also aids in the detection of irrelevant features, reducing overfitting and potentially improving performance. Furthermore, when a model contains minimal variables and reduces the complexity of model . It is the practice of picking the subset of most accurate and helpful characteristics to be used in model construction, either automatically or manually

Figure 2. Subset Selection

There are two types of feature selection models: supervised and unsupervised models. Filter, wrapper, and embedding methods are three different forms of supervised models. Filter techniques are applied in this case. In this way, features are deleted based on their relation to the output, or even how they relate to the outcome [17]. If the features are positively or adversely related to the output labels, we use correlation to assess if they should be dropped. From a list of attributes in a dataset, a subset of features from latitude, longitude, maximum temperature, minimum temperature, apparent temperature, humidity, pressure, and NDVI is selected for crop yield. For crop recommendation N, P, K, and rainfall and pH values are considered, as shown in Figure 3.

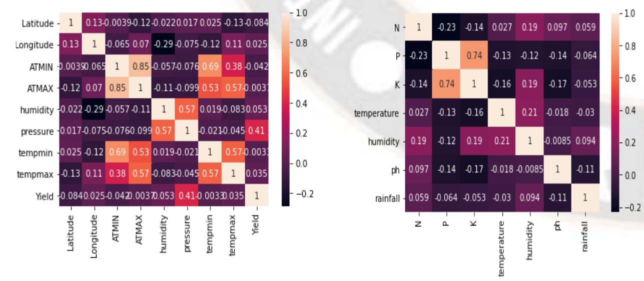

Figure 3. Features for crop yield and recommendation

b) *Model Building*

A sequential model (as shown in Figure 4) is created and trained for predicting crop yield. By building an object of the sequential class and applying model layers to it, one may construct deep-learning models with the sequential model API. The add() method is used to add layers to the neural network. To remove layers, use the pop() method. The sequential function Object() is used to generate a sequential model by supplying a list of levels to it. It is quite simple to determine the optimal number of layers. Continue to add layers until the test error no longer improves.

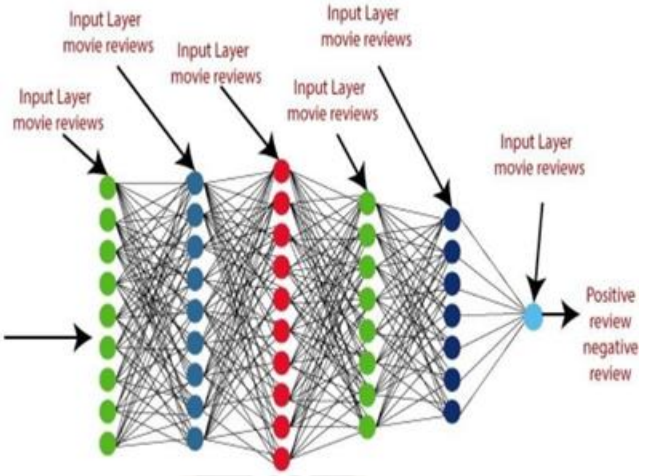

Figure 4. Sequential Model

A suitable activation function is selected for the model. Because neural networks are trained on large sets of data, the activation function should be effective and should reduce calculation time.

The activation function ReLU is utilized in this case. The ReLU or rectified linear activation function is a linear function that, if the input is positive, produces the same input as output; else, it generates zero. Because a network that uses it is faster to train and produces better outcomes, it can be stated numerically as:

$$R(z) = \max(0,z) \tag{1}$$

The ReLU function is defined as follows:

```
def ReLU(z):
if z>0:
return z
else:
return 0
```

The learning rate is taken into account during the learning process. To display the content of a model, execute the summary () method once it has been constructed.

c) *Train the model and predict crop yield*

A sequential technique is acceptable for a simple sequence of layers with one input tensor and one output tensor for each layer. The stochastic gradient descent-optimization approach is often used to train deep learning neural network models. An optimizer is a program or technique which alters features of a neural network,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

including the learning rate, and weights. In the end, it helps decrease overall loss and enhance accuracy. Instead of using the original data for every cycle, we use a random sample of data batches in stochastic gradient descent.

The very first phase of the process is to choose key parameters and learning rates [18]. Then, in every cycle, mix the data at random to reach an estimated minimum. The learning rate schedule is often used to control how the optimizer's learning rate varies with time. Rather than computing the aggregate of gradients of all instances, Stochastic Gradient Descent (SGD) calculates the gradient of a particular example's cost function at every cycle. SGD's path to the minima is frequently noisier than your standard method because just one sample from the data is selected randomly.

Figure 5 represents the block diagram of the model that is proposed. It contains all the steps involved in creating and training the model, successfully forecasting the crop yield, and also recommends the crop based on the given factors like N, P, K, humidity, and rainfall using various algorithms.

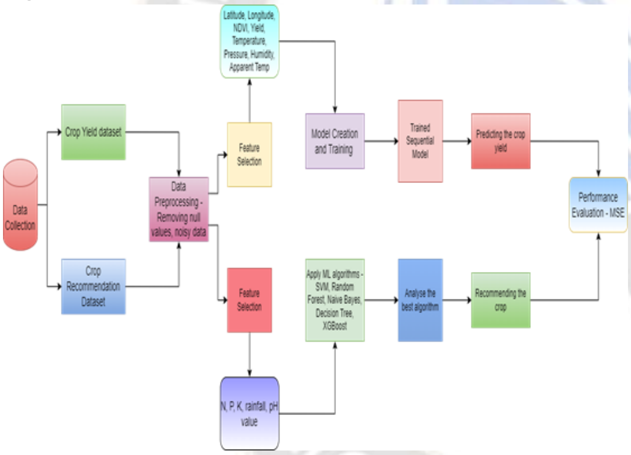

Figure 5. Block diagram of the proposed model

For training a model, a typically fit function is used. It uses a set number of epochs to train the model

Figure 6 represents the training sequential model.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 30)                270
_________________________________________________________________
dense_1 (Dense)              (None, 30)                930
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 31
=================================================================
Total params: 1,231
Trainable params: 1,231
Non-trainable params: 0
```

Figure 6. Training of Sequential Model

A webpage called CROPUP was developed to enter the values of latitude, longitude, apparent temperature, maximum and minimum temperature, pressure, humidity, and NDVI values which are most suitable to contribute to the yield of a crop (as shown in Figure 11). Based on the inputs given by the user, a sequential model with an SGD optimizer is employed to forecast crop yield. An optimizer is used to reduce the loss and improve the accuracy of the model. The 'predict' function is used in conjunction with the model class to provide output predictions for input samples in a batch manner.

In the case of DNN, the error function E is employed to calculate training accuracy. This procedure of learning is reiterated till DNN evaluates the association between the input, and the retrieved data, producing a group of trained categories that form the basis of the prediction.

$$E = 1/2\,(t-y)^2 \quad (2)$$

where t is the target output and y is the actual output of the neuron.

d) *Machine learning techniques*

1) DECISION TREE

The decision tree algorithm iteratively divides the overall sample space into smaller sub-sample spaces that can be articulated by a basic model. The entire sample space is held by the root node in the tree. Dividing a sample space into different subspaces involves dividing the root node into child nodes, with each child node being divided into leaf nodes repeatedly. That's a visual illustration for obtaining all feasible solutions to the issue depending on certain parameters. We employ the CART technique, which refers to the Classification and Regression Tree algorithm, to form a tree.

2) NAIVEBAYES

The Bayes theorem-based Naive Bayes technique is one of the simplest and most successful methods for generating fast predictive models which can generate fast predictions [19]. This is a classification algorithm, which implies that it makes predictions based on an entity's probability. We would fit the Naive Bayes classifier to training data after pre-processing data. To adapt it to the training data, a Gaussian NB classifier is usually applied. Different classifiers could also be used depending on our needs. The Naive Bayes Classification algorithm uses the Bayes rule to calculate the conditional probability of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

an event A provided an event B. It is straightforward to perform. It doesn't need quite so much data for training.

3) SUOOORT VECTOR MACHINE(SVM)

SVM is employed to resolve both regression and classification issues. The SVM method aimed to figure out the finest line as well as decision boundary for classifying n-dimensional space such that extra data points could be conveniently inserted in the right group hereafter. A hyperplane is a boundary that represents the optimum decision. Utilizing kernel functions including linear, polynomial, and others, this algorithm finds the best differentiating classifier between two classes by increasing the margin to the max between support vectors. The purpose of the support

4) LOGIDTIC REGRESSION

Logistic regression can be used to classify observations based on a variety of characteristics; it can immediately detect the most useful variables for classification. It is a form of regression that employs the theory of predictive analysis. Rather than fitting a regression line, we fit an "S" structured logistic function in logistic regression, that forecasts 2 max values that are 0 and 1. The sigmoid function is a mathematical equation for transforming expected values into probabilities.

5) RANDOM FOREST

Random Forest is a type of machine-learning technique that can be employed to solve regression and classification problems. It relies on ensemble methods, which is a method of combining numerous classifiers to solve complex problems and improve the model's performance. Multiple techniques of the same kind are merged in the random forest classifier. The value of the average prediction of different trees is often used to evaluate the forest's output. The random tree is identical to C4.5 or CART, except that it simply picks an arbitrary subset of features. The higher the density of trees in the forest, the greater the efficiency, and the danger of overfitting is averted. Compared to other approaches, it requires lesser time for training.

6) XGBOOST

XGBoost is a technique of ensemble model. Ensemble learning is a method for combining the prediction performance of several learners in a structured way. Ensemble modeling is a useful tool for improving model's performance. The ensemble's foundation learners, or models, maybe from the same or various learning algorithms. In XGBoost, gradient boosted decision trees are developed. Weights are assigned to each of the independent factors, which are then fed into the decision tree, that anticipates outcomes.

e) Finding the best algorithm to recommend crop

The trained model's parameters are evaluated, and effective predictive values are provided. Metrics like root mean square error, mean squared error, mean absolute percentage error, and mean absolute error can be used to evaluate the performance of the model. The loss and accuracy of a trained model is assessed using the mean squared error (MSE) measure. In the recommendation system, the accuracy of 99.31% is obtained

XGBoost is better when compared to other algorithms in giving importance to functional space and reduces the cost of model. Table 3 represents the accuracy of various algorithms used for crop recommendation. Figure 7 represents the yield of the crop with respect to pressure.

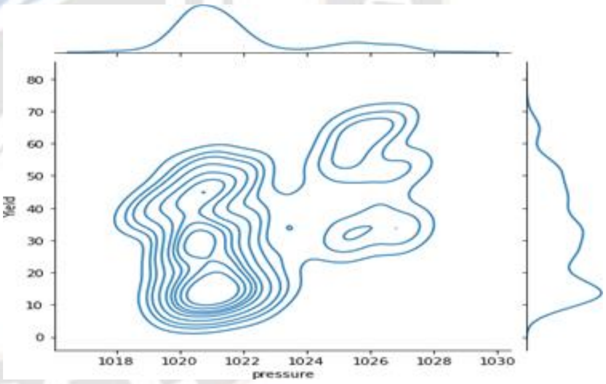

Figure 7. Figure 7. Joint-plot of yield and pressure

Figure 8 represents the yield obtained concerning count, which is obtained using a count plot.

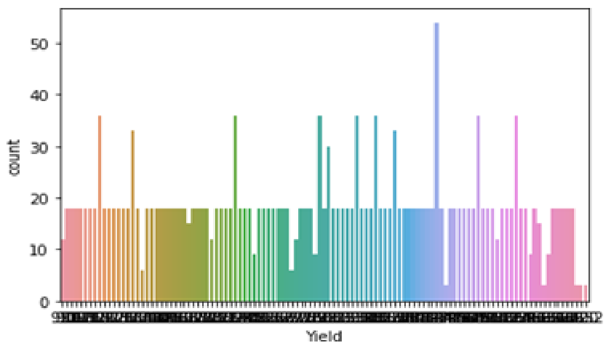

Figure 8. Count-plot for crop yield

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

TABLE III. ACCURACY OF VARIOUS MACHINE LEARNING ALGORITHMS USED FOR CROP RECOMMENDATION SYSTEMS.

Algorithm	Accuracy
DecisionTree	90%
NaïveBayes	99.09%
SupportVectorMachine	97.95%
LinearRegression	95.22%
RandomForest	99.09%
xgboost	99.31%

Figure 9 illustrates that accuracy obtained using XgBoost is greater when compared to various techniques of machine learning like SVM, Decision Tree, Linear Regression, Naïve Bayes, RF, and XgBoost.

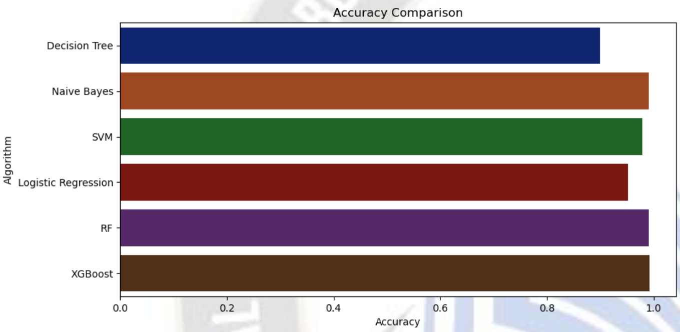

Figure 9. Accuracy of various algorithms for crop recommendation system

IV. EXPERIMENTAL RESULTS

The algorithm, which has been trained, can accurately predict crop production and recommend the best crop. Throughout India, the model built has a 99.3 percent accuracy rate for crops in all seasons. We can determine the agricultural output based on numerous factors using a Deep Neural Network sequential model. The input parameters that are used by our model are latitude, longitude, apparent temperature, pressure, humidity, temperature maximum, temperature minimum, and NDVI entered by the user in a web page interface that depicts the crop's predicted yield. in kilograms (kg) per hectare (as shown in Figure 12). A crop recommender suggests a suitable crop based on the XGBoost technique. Farmers can benefit from the proposed approach even if they just have a small agricultural area by knowing about the yield of a crop and recommended crops to be grown in the field which can help maximize the yield (as shown in Figure 13, 14, 15, 16).

Figure 10. Home Page

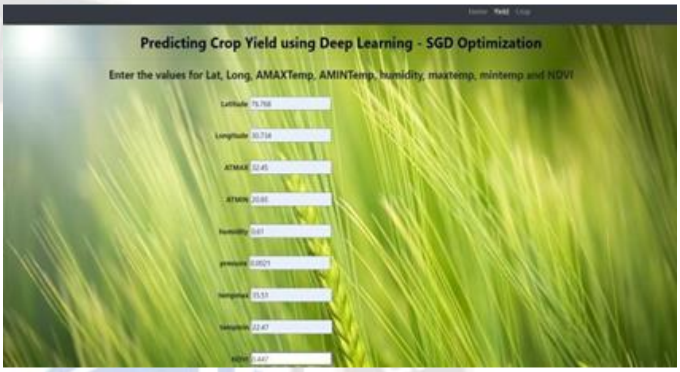

Figure 11. Enter the values to predict the Yield

Figure 12. Predicted Yield in Kg/ha

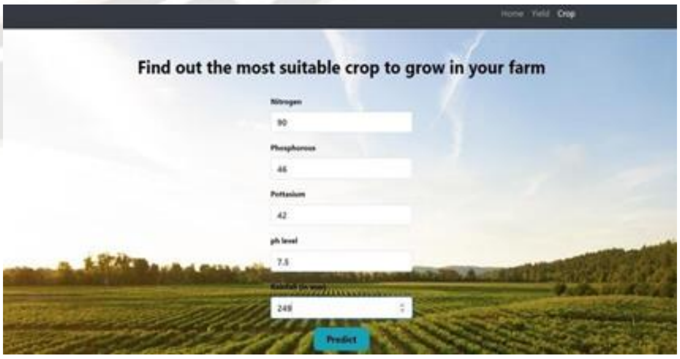

Figure 13. Enter the values to recommend suitable crop

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

Figure 14. The recommended crop is rice

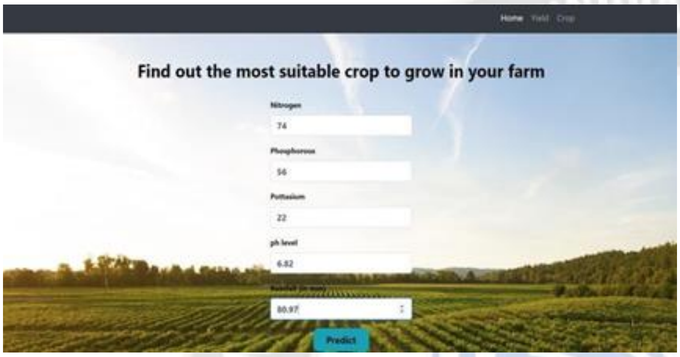

Figure 15. Enter another set of values

Figure 16. The recommended crop is maize

V. Conclusion and Future Outlook

In this study, we have proposed a neural network model for predicting crop yields . It is based on textual data . one can use it to determine what crop type is ideal for their conditions and get crop recommendations. Crop recommendation is also a feature of this system that helps farmers to decide on the most suitable crop to grown on their farm. Farmers can use a crop production model to help them select what to grow and when to plant it. It acts as a support system for the farmer to help him to improve his farming activities and maximize productivity. The focus of future effort will be on implementing a system, following identification of nutritional deficits in the soil by analyzing nitrogen, potassium, phosphorus, etc. we would be able to suggest fertilizer to farmers. We also plan on including disease detection along with the fertilizer suggestions so that farmers can detect the disease of the crop ahead of time and take the appropriate measures.

References

[1] Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access, 9, 63406-63439. doi: 10.1109/ACCESS.2021.3075159

[2] Kalaiarasi, E., & Anbarasi, A. (2021). Crop yield prediction using multi-parametric deep neuralnetworks. *Indian JournalofScienceand Technology*, *14*(2),131-140.

[3] Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers inplantscience*,*10*, 621. doi: 10.3389/fpls.2019.00621

[4] Kale, S. S., & Patil, P. S. (2019, December). A machine learning approach to predict crop yieldand success rate. In *IEEE Pune Section International Conference (PuneCon)* (pp. 1-5). IEEE.doi: 10.1109/PuneCon46936.2019.9105741

[5] Dahikar, S. S., & Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. International journal of innovative research in electrical, electronics, instrumentation and control engineering, 2(1), 683-686.

[6] Ren, J., Chen, Z., Yang, X., Liu, X., & Zhou, Q. (2009, July). Regional yield prediction of winter wheat based on retrieval of Leaf area index by remote sensing technology. In 2009 IEEE International Geoscience and Remote Sensing Symposium (Vol. 4, pp. IV-374). IEEE.doi: 10.1109/IGARSS.2009.5417391

[7] Chuan, J., Qiming, Q., Lin, Z., Peng, N., & Ghulam, A. (2007, July). TVDI based crop yield prediction model for stressed surfaces. In 2007 IEEE International Geoscience and Remote Sensing Symposium (pp. 4656-4658). IEEE. doi: 10.1109/IGARSS.2007.4423896Y.

[8] Mythili, K., & Rangaraj, R. (2021). Deep Learning with Particle Swarm Based Hyper Parameter Tuning Based Crop Recommendation for Better Crop Yield for Precision Agriculture. Indian Journal of Science and Technology, 14(17), 1325-1337. doi: 10.17485/IJST/v14i17.450M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[9] Pavani, S., & Beulet, A. S. (2019). Heuristic prediction of crop yield using machine learning technique. International Journal of Engineering and Advanced Technology (IJEAT), 135-138.

[10] Bondre, D. A., & Mahagaonkar, S. (2019). Prediction of crop yield and fertilizer recommendation using machine learning algorithms. International Journal of Engineering Applied Sciences and Technology, 4(5), 371-376.

[11] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015, May). Crop Selection Method to maximize crop yield rate using machine learning technique. In 2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM) (pp. 138-145). IEEE. doi: 10.1109/ICSTM.2015.7225403

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
DOI: https://doi.org/10.17762/ijritcc.v10i11.5780
***Article Received**: 27August 2022 **Revised**: 05October 2022 **Accepted**: 20October 2022*

[12] Shanmuganathan, S., Sallis, P., & Narayanan, A. (2010, July). Data Mining Techniques for Modelling the Influence of Daily Extreme Weather Conditions on Grapevine, Wine Quality and Perennial Crop Yield. In 2010 2nd International Conference on Computational Intelligence, Communication Systems and Networks (pp. 90-95). IEEE. doi: 10.1109/CICSyN.2010.15

[13] Suresh, G., Kumar, A. S., Lekashri, S., & Manikandan, R. (2021). Efficient crop yield recommendation system using machine learning for digital farming. International Journal of Modern Agriculture, 10(1),906-914.

[14] Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters, 14(5), 778-782. doi: 10.1109/LGRS.2017.2681128

[15] Oikonomidis, A., Catal, C., & Kassahun, A. (2022). Deep learning for crop yield prediction: a systematic literature review. New Zealand Journal of Crop and Horticultural Science,1-26.doi: https://doi.org/10.1080/01140671.2022.2032213

[16] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: Intelligent crop recommendation system using machine learning algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE. doi: 10.1109/ICCUBEA.2018.8697349

[17] Kuwata, K., & Shibasaki, R. (2015, July). Estimating crop yields with deep learning and remotely sensed data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp.858-861).IEEE. doi: 10.1109/IGARSS.2015.7325900

[18] Amaratunga, V., Wickramasinghe, L., Perera, A., Jayasinghe, J., & Rathnayake, U. (2020). Artificial neural network to estimate the paddy yield prediction using climatic data. Mathematical Problems in Engineering, 2020. doi: 10.1155/2020/8627824

[19] Gandhi, N., Armstrong, L. J., & Petkar, O. (2016, September). PredictingRice crop yield using Bayesian networks. In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 795-799). IEEE.doi: 10.1109/ICACCI.2016.7732143