Indian Institute of Information Technology Pune
Department of Computer Science Engineering 2022-2023

Project on
"Smart Farming Crop Yield Prediction using Machine Learning"

V. Bharadwaj - (112015160)
G.Manoj Reddy - (112015054)
R.Suraj Rao - (112015108)
U.Vinod Kumar - (112015158)

Project Supervisor - Prof. Ritu Tiwari

Outline

- Introduction
- Motivation
- Objective
- Literature Survey
- Methodology
- Results and Discussion
- Conclusion and future scope
- References

Introduction

Agriculture is the backbone of Indian economy. Farming is a highly difficult and demanding occupation. Due to a variety of causes, farmers in our nation endure great challenges. Lack of knowledge or inaccurate information is the major cause of their suffering. By ensuring that our cultivators receive more sophisticated information about the best crop that can be harvested to ensure maximum yield by taking into account factors like humidity, rainfall, soil nutrition, and precipitation levels in their locality, the agriculture sector in our nation could be significantly strengthened. We can provide farmers the most accurate forecasts and recommendations via the use of contemporary machine learning algorithms, enabling them to generate high-quality yields and enjoy the profits.

Motivation

- Farmers may only benefit from the ultimate yield if they harvest the crop that is appropriate for the field's circumstances. This may be achieved by carefully examining the soil and climate data.

- We require a lot of soil content and nature estimations and predictions. Machine learning and deep learning algorithms may be used to easily make these predictions based on vast datasets.

Objective

The main objectives of this project are:

- To increase the accuracy of crop yield prediction by using different machine learning and deep learning algorithms.
- To help farmers to choose a suitable crop to cultivate in order to maximize the yield by predicting the disease with image detection.
- Also Recommends Farmers to use right Pesticides for the diseases.
- To give farmers a generic idea on how external factors (like temperature,rainfall etc.)can influence their yield.

Literature Survey

Author	Title	Learnings
Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Vinu Williams	Crop Yield Prediction using Machine Learning Algorithms,2021	This Paper is mainly focused on using machine learning algorithms like Logistic regression, Naive bayes, Random forest.
Dhivya Elavarasan, P. M. Durairaj Vincent	“Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications”, IEEE Access, 2020	Focuses on crop yield prediction using deep reinforcement learning techniques like ANN and RNN
Thomas van Klompenburg, Ayalew Kassahun,Cagatay Catal	“Crop yield prediction using machine learning: A systematic literature review” Computers and Electronics in Agriculture, Volume 177, October 2020	This paper is mainly focused on ML algorithms like linear regression,random forest,gradient boosting tree and DL algorithms like CNN , DNN and LSTM.

Author	Title	Learnings
Sangeeta, Shruthi G	“Design And Implementation Of Crop Yield Prediction Model In Agriculture”,2020	This Paper includes Machine learning algorithms like random forest,polynomial regression,decision trees for forecasting crop yield based on climatic parameters
YongKang Xing, JiaPeng Huang, YongYao Lai	“Research and Analysis of the Front-end Frameworks and Libraries in E-Business Development”, Proceedings of the 2019 11th International Conference on Computer and Automation Engineering - ICCAE 2019, 2019.	This paper describes three different front-end development frameworks and libraries for developing web applications and also lists potential solutions for developing a web application. Uses React, Angular 2, Vue
Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa	“Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY”, 2019	

Methodology

Step1: Import datasets- Collect datasets and merge these datasets in a structured form.

Step 2: Data Preprocessing -It is done to remove inaccurate, incomplete and unreasonable data which results in increase in quality of the data and hence the overall productivity using standard preprocessing techniques.

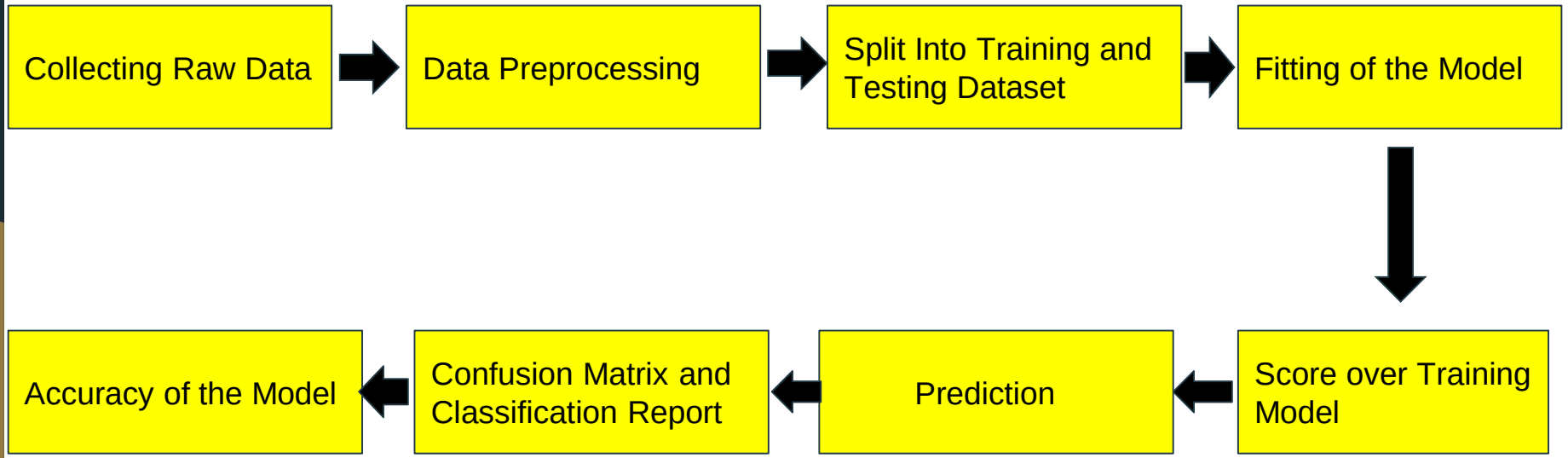
Step 3: Train-Test Splitting- Divide the analysed crop data into training and testing sets and train the model using the training data to predict the crop yield for given inputs.

Step 4: Feature Extraction- It is the process of reducing the raw data into manageable groups (features) for processing it. Beginning with an initial set of raw data it builds up derived values (features) which results in an informative and non-redundant data.

Step 5: Algorithms- Compare results of various algorithms by passing the analysed dataset through them and calculating the error rate and accuracy for each. Choose the machine learning algorithm with the highest accuracy and lowest error rate.

Step 6: Testing- Test the implemented system on the testing set to check for accuracy and failures.

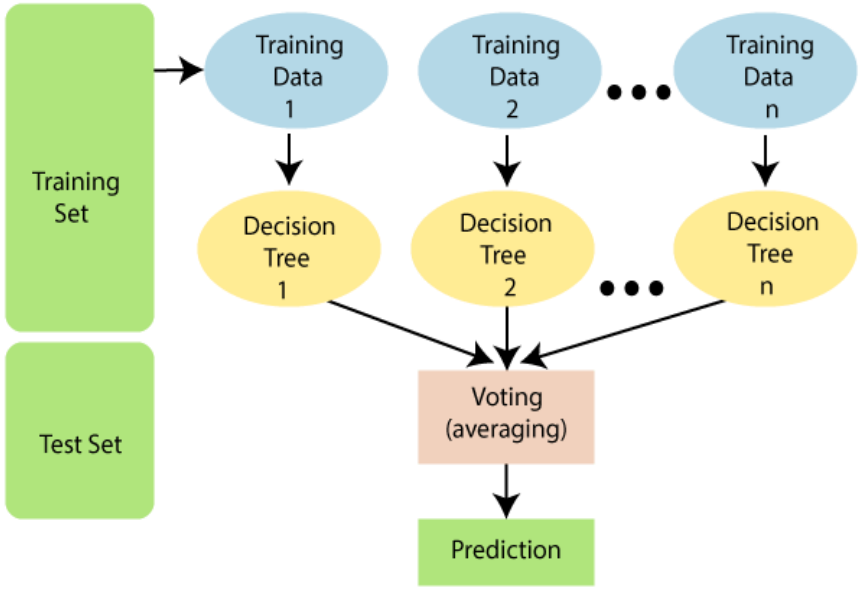
Visual Representation of Methodology



Random Forest Regression

Random forest regression is an ensemble machine learning technique.

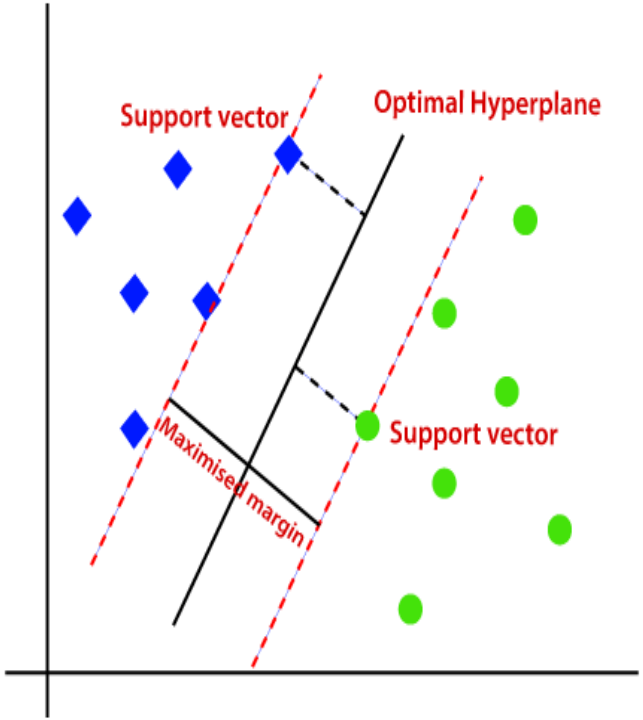
- A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.



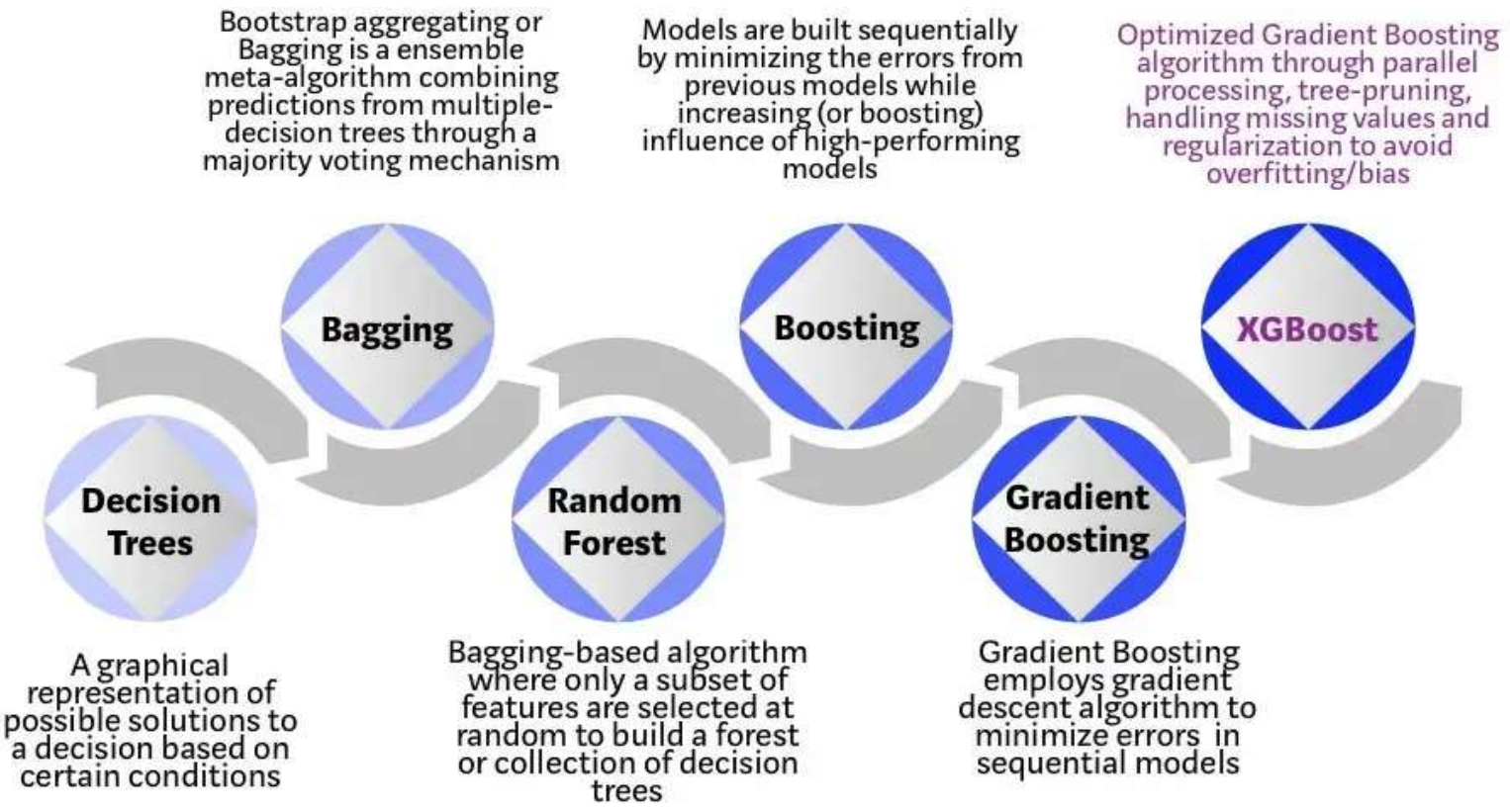
Support Vector Regression

Support Vector Regression(SVR) uses the Support Vector Machine(SVM) algorithm to predict a continuous variable.Support Vector Regression tries to fit the best line within a predefined or threshold error value.

- Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
- In Support Vector Regression, the straight line that is required to fit the data is referred to as hyperplane.
- This algorithm acknowledges the presence of non-linearity in the data and provides a proficient prediction model.

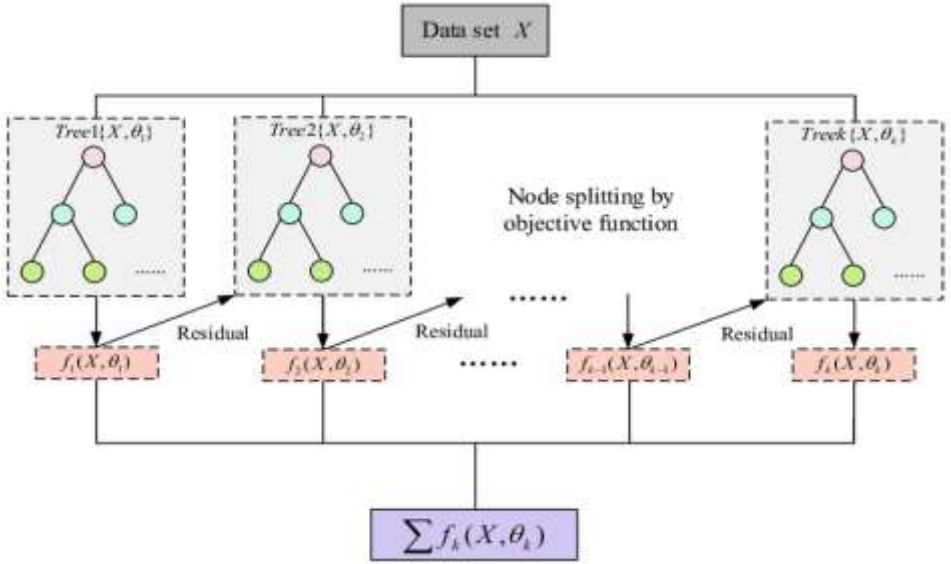


XGBoost


Bootstrap aggregating or Bagging is a ensemble meta-algorithm combining predictions from multiple-decision trees through a majority voting mechanism
Models are built sequentially by minimizing the errors from previous models while increasing (or boosting) influence of high-performing models
Optimized Gradient Boosting algorithm through parallel processing, tree-pruning, handling missing values and regularization to avoid overfitting/bias
Bagging
Boosting
XGBoost
Decision Trees
Random Forest
Gradient Boosting
A graphical representation of possible solutions to a decision based on certain conditions
Bagging-based algorithm where only a subset of features are selected at random to build a forest or collection of decision trees
Gradient Boosting employs gradient descent algorithm to minimize errors in sequential models

- XGBoost is an implementation of Gradient Boosted decision trees. In this algorithm, decision trees are created in sequential form.
- Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. The weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree.

- These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression, classification, ranking, and user-defined prediction problems.



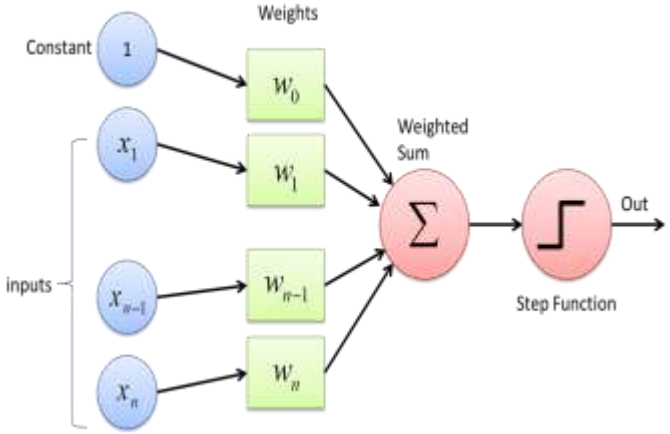
Artificial Neural Networks

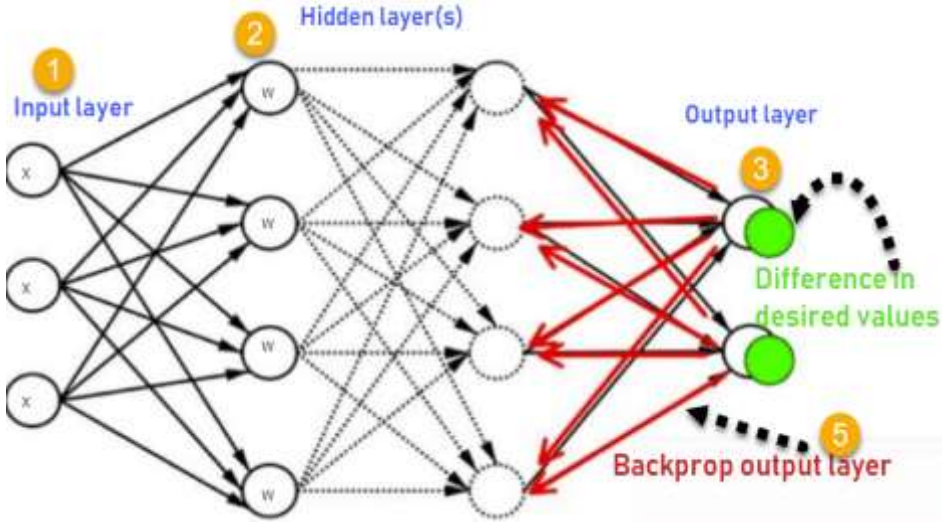
- Artificial Neural Network can be best represented as a weighted directed graph, where the artificial neurons form the nodes.
- Artificial Neural Network primarily consists of three layers: Input layer, Hidden layer , Output layer
- The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function. It determines weighted total is passed as an input to an activation function to produce the output.
- Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.

1. Inputs X, arrive through the preconnected path
2. Input is modeled using real weights W. The weights are usually randomly selected.
3. Calculate the output for every neuron from the input layer, to the hidden layers, to the output layer.
4. Calculate the error in the outputs

   Error(B) = Actual Output – Desired Output

5. Travel back from the output layer to the hidden layer to adjust the weights such that the error is decreased.





Results and Discussion

- The various machine learning algorithms applied in the project gave different kinds of results with various accuracy scores.
- These algorithms have been compared based on their mean squared error(MSE), mean average error(MAE) and R2 score.
- The Mean Squared Error measures how close a regression line is to a set of data points. It is a risk function corresponding to the expected value of the squared error loss.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$

- Mean Absolute Error calculates the average difference between the calculated values and actual values

$$\mathbf{MAE = (1/n) * \sum |yi - xi|}$$

- It is the amount of the variation in the output dependent attribute which is predictable from the input independent variable(s)

$$\mathbf{R2 = 1 - SSres / SStot}$$

Results of Random Forest Algorithm

- The following are the results for the Random Forest Algorithm:

Result Type	Value
Mean Squared Error(MSE)	7.671048879964048
Mean Average Error(MAE)	0.8953650873829122
R2 Score	0.9589614680509005

Results of Support Vector Machine

- The following are the results for the Polynomial Support Vector Machine(SVM) Algorithm:

Result Type	**Value**
Polynomial SVM Score	0.6312485850825429
R2 Score	0.9598856437260073

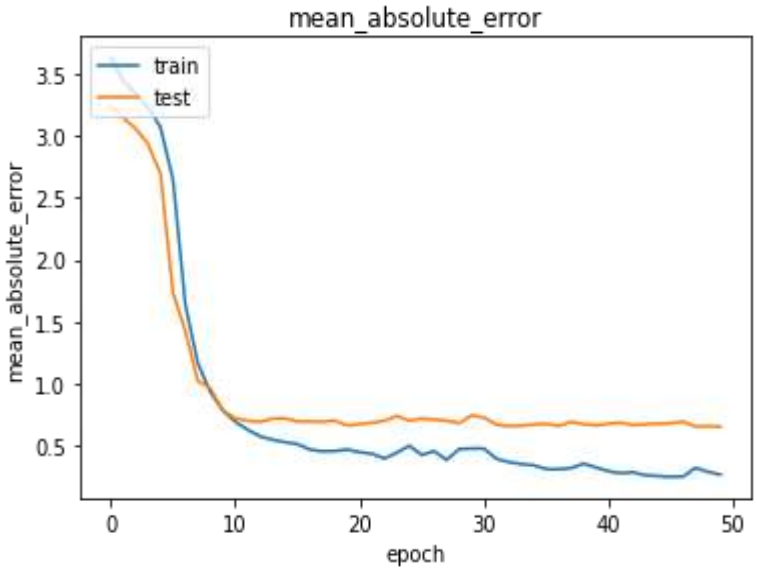
Results of XGBooster Algorithm

- The following are the results for the XGBooster Algorithm:

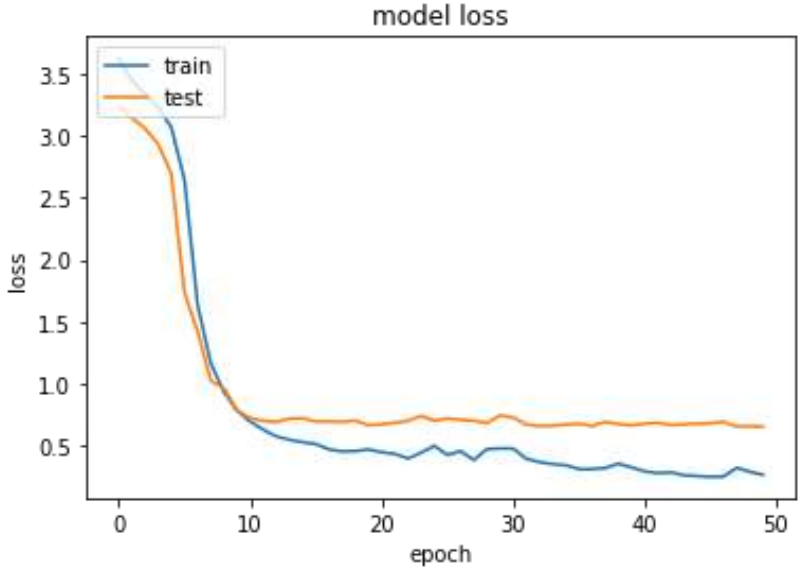
Result Type	Value
XGBoost validation MAE	0.6670485576475581
R2 Score	0.9654928330252374

Results of Neural Network algorithms

- The following are the results of the Neural Network Algorithm:



Mean Absolute Error



Model loss

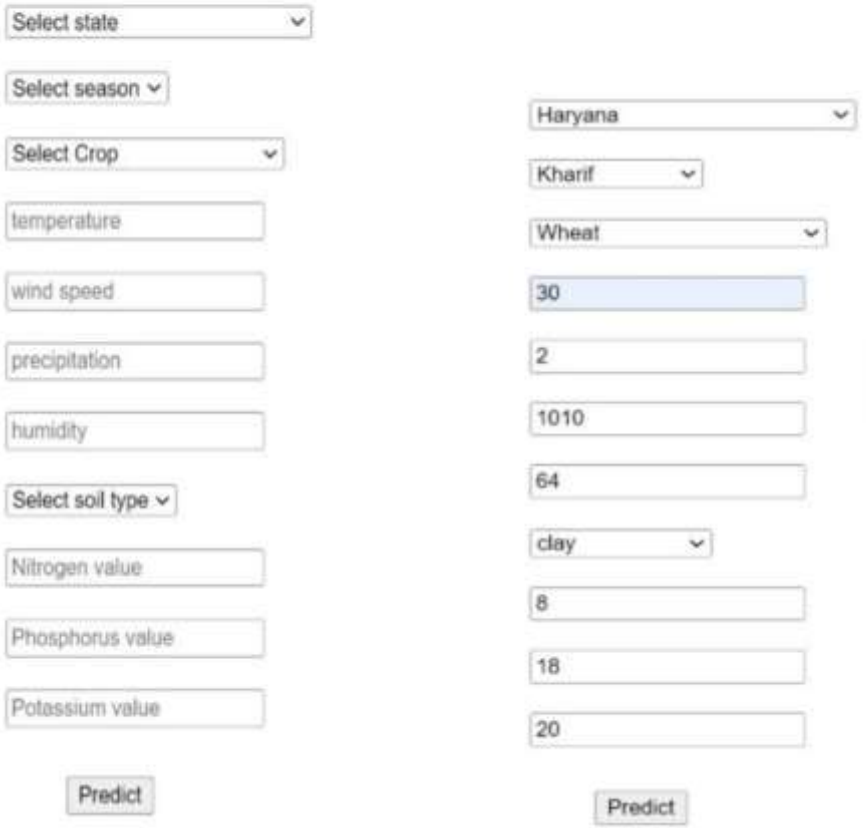
Code for website

predict.html | app.py | index.html | complete_dataset.csv

app.py > ...

```python
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle

# Create flask app
flask_app = Flask(__name__)
model = pickle.load(open("xgboost_yield_prediction_final.pkl", "rb"))

@flask_app.route("/")
def Home():
    return render_template("index.html")

@flask_app.route('/predict', methods=['POST'])
def predict():

    float_features = [float(x) for x in request.form.values()]
    features = [np.array(float_features)]
    final_prediction = model.predict(features)
    result_str = ''.join(str(e) for e in final_prediction)
    final_prediction=result_str
    return render_template('predict.html', prediction_text=final_prediction)



if __name__ == "__main__":
    flask_app.run(debug=True)
```

Crop Yield Prediction



Output:-

Crop Yields *4.0793695* quintal for an acre

Discussion of Results

- From the given values of errors,we can understand that Extreme Gradient Booster(XGBooster) is the algorithm which gives the highest R2 Score.This is because in XGBooster, after the model is trained for a part of the dataset,then the model is tested. After the testing is done,the weights of the misclassified points are increased so the probability of selecting them is increased.In this way,we ensure that the misclassified points get trained more so that the machine learning algorithm can classify those points correctly in future.
- Because of this regular updation of weights and repeated training, XGBooster algorithm gives the highest accuracy while predicting.

Conclusion and Future Scope

- Our system helps farmers predict the yield of a given crop and also helps them to decide which crop to grow. The system also provides a list of crops with their productions based on climatic conditions.
- One of the most important and novel contributions of the system is to suggest to the user the right time to apply fertilizer, by forecasting the weather for the next 14 days.

- The Future Work focuses on a fully automated system that will provide a sequence of crops to grow depending on soil and weather conditions and update the datasets from time to time to make accurate predictions.
- In the future, this model can be implemented across India by adding data points for all regions. Our system can be integrated with a messaging module so that registered farmers can receive forecast notifications directly to their registered mobile numbers.

References

1. Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Vinu Williams Crop Yield Prediction using Machine Learning Algorithms, 2021
2. Dhivya Elavarasan, P. M. Durairaj Vincent "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications", IEEE Access, 2020.
3. Thomas van Klompenburg, Ayalew Kassahun,Cagatay Catal "Crop yield prediction using machine learning: A systematic literature review" Computers and Electronics in Agriculture, Volume 177, October 2020.
4. Sangeeta, Shruthi G, "Design And Implementation Of Crop Yield Prediction Model In Agriculture",2020.
5. YongKang Xing, JiaPeng Huang, YongYao Lai "Research and Analysis of the Front-end Frameworks and Libraries in E-Business Development", Proceedings of the 2019 11th International Conference on Computer and Automation Engineering - ICCAE 2019, 2019.
6. Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa,"Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY", 2019.
7. P.Priya, U.Muthaiah & M.Balamurugan Predicting Yield of the Crop Using Machine Learning Algorithm, International Journal of Engineering Sciences & Research Technology (IJESRT), April, 2018.
8. S. Pavani, Augusta Sophy Beulet P "Heuristic Prediction of Crop Yield using Machine Learning Technique", International Journal of Engineering and Advanced Technology (IJEAT) Volume-9, December 2019.
9. Sriram Rakshith.K,Dr. Deepak.G, Rajesh M, Sudarshan K S, Vasanth S & Harish Kumar, "A Survey on Crop Prediction using Machine Learning Approach" International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 7, Issue IV, Apr 2019