CONTENTS

1. Abstract

Agriculture is a vital industry in the development of our country's economy. Agriculture is what gave rise to civilization. India is an agrarian country with an economy focused primarily on crop productivity. As a result, we may claim that agriculture is the foundation of all business in our country. The selection of each crop is critical in agricultural planning. Crop selection will be influenced by various factors such as market price, production rate, and government policies. Many adjustments in agriculture are essential to improve changes in our Indian economy. We can improve agriculture by utilizing machine learning techniques that are easily used in the farming sector.

Along with advancements in farming tools and technologies, valuable and reliable knowledge on many topics is also important. The goal of this research is to develop an appropriate machine learning model to estimate crop yield, so that this technology can aid in the resolution of various agricultural and farmer problems. This benefits our Indian economy by increasing agriculture production yield rates.

Keywords: Crop Yield, Machine Learning techniques, Crop yield prediction, Indian Agriculture

2. Introduction:

Agriculture has an important role in everyone's lives. Agriculture has been regarded as one of the most important practices in India from ancient times. People used to cultivate crops on their own property to suit their needs in the past. As India's backbone, the agricultural sector has grown in response to public demand as technology advances. With fast population expansion, these advances are critical to meeting the requirements of everyone.

In the previous few years, the price of onions in our country has fluctuated significantly. As a result of the big increase in price, most farmers opted to plant onion in their fields in order to generate large profits from their land. This resulted in an abundant supply of onions in some locations, such as Maharashtra, while many other regions had crop failure and farmers lost a significant amount of money. This issue arose as a result of numerous unfavorable conditions that inhibited onion growth. A continued shortage of onions in the coming months had a devastating impact on the lives of ordinary people. This occurred because middle-class people were unable to afford the high price of onion, a commonly utilized product in their homes.

Machine learning (ML) methods are utilized in a variety of fields, from supermarkets to evaluate customer behavior to phone usage prediction. For many years, machine learning has been employed in agriculture. Crop production prediction is one of the most difficult challenges in precision agriculture, with several models presented and proven thus far. This challenge necessitates the use of many datasets because agricultural output is affected by numerous factors such as climate, weather, soil, fertilizer use, and seed variety.

This suggests that crop production prediction is not a simple operation, but rather a series of complex steps. Crop yield prediction methods may now reasonably approximate actual yield, but improved yield prediction performance is still desired. Machine learning, a subset of Artificial Intelligence (AI) that focuses on learning, is a practical approach that can provide superior yield prediction based on a variety of characteristics.

Machine learning (ML) can identify patterns and correlations in datasets and uncover information. The models must be trained using datasets that describe the outcomes based on previous experience. The predictive model is developed with numerous features, and as a result, model parameters are calculated using previous data during the training phase. Part of the previous data that was not used for training is used for performance evaluation during the testing phase. Depending on the study challenge and research objectives, an ML model might be descriptive or predictive.

The preceding example helps us formulate our problem statement.

Problem Statement

India is mostly an agricultural country. Each crop's yield is determined by its dependent elements. It is critical to predict agricultural yield in order to assist farmers. Crop Yield Prediction is the prediction of a crop's yield in the future based on the dependant elements. Crop yield is affected by elements such as rainfall, pressure, temperature, pesticide application, and so on. This is accomplished by developing a system for predicting crop yield.

Relevance of the Project

This project is built on alleviating farmers' problems as well as offering an actual level of harvest they may expect from the crop they have grown based on dependant factors such as temperature, rainfall, and so on. This project was created primarily to assist farmers in analyzing the yield of their crop. Farmers are losing agricultural yield owing to a lack of understanding about the crop and the natural factors affecting them. In this project, we examine the characteristics that predict crop production well before harvest.

Motivation

Farmers have reported difficulties with agricultural yield due to abrupt fluctuations in the weather, which have an impact on crop yield. Reduce crop quality, resulting in lower crop income for producers. This project aims to improve crop quality, which will allow farmers to earn more money. In this project, we collected datasets for all of the factors that are affected by crops across a number of years. The prediction is obtained using this data to show that the harvest of the crop that is growing in that region.

India is mostly an agricultural country. Each crop's yield is determined by its dependent elements. It is critical to predict agricultural yield in order to assist farmers. Crop Yield Prediction is the prediction of a crop's yield in the future based on the dependant elements. Crop yield is affected by elements such as rainfall, pressure, temperature, and geographical location. Machine learning algorithms are used to predict dependent elements based on previous data.

Based on the algorithmic forecast, this Crop Yield Prediction model will assist farmers in growing crops that provide higher yields. As a result, it helps farmers cut their losses. Prediction also aids in the growth of the national economy.

Scope

We present a scalable, accurate, and low-cost technique for predicting agricultural production using existing climate data and machine learning in this research. Our machine learning method can accurately estimate crop yield (for particular crop). We obtained temperature, rainfall, crop yield, and other datasets from numerous sources in India, including the Indian Methodological Department and the Agriculture Department. Machine learning methods such as Extra Trees Regressor, multi-linear regression, clustering, SVM, and others can be used to estimate crop yield based on variables such as temperature, rainfall, and pressure.

Existing works

Many machine learning techniques are used in agriculture to increase crop yield rates. Crop productivity may be affected by regional geographical circumstances such as river bed, hill areas, or depth areas. Weather conditions such as humidity, rainfall, temperature, and cloud cover. Clay, sandy, saline, or peaty soils are all possible. Soil composition can include copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph value, or carbon, as well as various harvesting methods. Many parameters are utilized for various crops to make predictions. As a result, our farmers must be aware of any new machine learning and other techniques. These approaches aid in increasing crop productivity.

Dataset Used

- The chosen dataset was taken from Kaggle Repository.
- Contains the crop yield production of 10 crops from all states and union territories from India. Records start from 1990 to 2013.
- 20,847 Rows with 7 attributes: Area, Item, Year, Yield (hg/ha), Average Rainfall (mm), Pesticides, Average Temperature.
- The first few rows are as follows:

	Area	Item	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_to	avg_temp
1	Andaman & Nicobar Islands	Cassava	1990	31400	1996	12004.33	19.64
2	Andaman & Nicobar Islands	Maize	1990	20052	1996	12004.33	19.64
3	Andaman & Nicobar Islands	Plantains and oth	1990	394286	1996	12004.33	19.64
4	Andaman & Nicobar Islands	Potatoes	1990	210685	1996	12004.33	19.64
5	Andaman & Nicobar Islands	Rice, paddy	1990	31111	1996	12004.33	19.64

3. Work/Literature Survey:

S.no	Author	Conclusion
1.	Explainability Analysis of a Machine Learning Model for Industrial Applications. Authors:Ramon Gomes Durries, Turıbio Tanus Salis, Frederico Gualberto Ferreira Coelho, Antonio de Padua Braga DOI:10.1109/ICECET55527.2022.9872890 (2022)	• In this work, the authors have trained a machine learning model to predict the yield strength of seamless steel tubes at the end of the heat treatment process. They then perform a detailed analysis of theexplainability charts generated by the TreeSHAP (SHapleyAdditiveexPlanations) method for the model. • Pros:Their analysis validates the model, showing that its predictions are based on expected relations between input features and their values. It also allowed for the extraction of nontrivial knowledge regarding the process and the model itself. • Cons: Despite being widespread in the academia, the adoption of such technologies in the manufacturing industry is slow.
2.	A Review on Prediction of Crop Yield usingMachine Learning Techniques. Authors:Mr. Sunil G L, Dr.Nagaveni V, Mrs. Shruthi U DOI:10.1109/TENSYMP54529.2022.9864482 (2022)	• The authors have carried out acomparative study on supervised and unsupervised machinelearning techniques for prediction of crop yield. • Pros:In thesupervised learning method, the classification method achieved better results. In the Unsupervised method technique, clustering was seen to be the better option. • Cons: The classification method required a lot more parameters than the regression model. And although clustering was found to be a viable option, we can obtain more accuracy with Bi-Clustering as it can solve more complex problems.
3.	Integrating Crop Simulation and Machine LearningModels to Improve Crop Yield Prediction. Authors:Saiara Samira Sajid, Isaiah Huber, SotiriosArchontoulis, Guiping Hu	• This research focuses on identifying a robust crop prediction model that can be applied on a large scale. To build the prediction model results from crop simulation software(APSIM) were combined with ML models. Finally, we developed an optimized weighted ensemble model. In this model, optimum weights were assigned to the prediction results from base

	DOI:10.1109/SOSE55472.2022.9812678 (2022)	models. • Pros:Comparing the results,our proposed model obtained an RRMSE of 8.99%, while the black-box method had an RRMSE of around 8.5%. Hence, theproposed model can maintain similar accuracy as the CNN-DNN model while with additional benefits in interpretability. Moreover, the optimum weights were calculated by minimizing the RMSE of the model, which allows the model to be applied on a larger scale without compromising efficiency. • Cons:by evaluating the correlation between weather and soil properties, it was found that thesource of error in the model is random. This is not true in the real world as soil and weather play an important role in crop yield.
4.	Crop yield prediction: two-tiered machine learning model approach Authors:SushilaShidnal, Mrityunjaya V Latte, Ayush Kapoor DOI:10.1007/s41870-019-00375-x (2019)	• In this paper the nutrient deficiency of a paddy crop is considered. Tensor Flow's (Google's Machine Learning Library) is used to build a neural network to classify them into nitrogen, potassium, phosphorous deficiencies or healthy independently. It is necessary to have an optimal balance between nitrogen, potassium and phosphorous content. • Results:Tensor Flow's model identifies the deficiency using a set of images. The result is fed to "machine learning driven layer" to estimate the level of deficiency on a quantitative basis. It specifically makes use of k means-clustering algorithm. It is then evaluated through the rule-matrix to estimate the cropland's yield. A fair prediction of 76–77% was observed with two tired machine learning models. • Future Scope: The accuracy of the prediction is low for a real-time, high volume streamlined application. Deep-learning techniques adversarial networks are being looked at for this application. Data collection is still a challenge.
5.	A Deep Network Solution for Attention and Aesthetics Aware Data Capturing Authors: W. Wang, J. Shen and H. Ling DOI: 10.1109/TPAMI.2018.2840724.	• Compared with previous arts, we treat the photo cropping task in a more natural and efficient way, with the following major contributions: A deep learning framework to combine attention, High computation efficiency, and aesthetics components for data capture. This is achieved by The proposed deep model is composed of two subnetworks: an Attention Box Prediction (ABP) network and

	(2022)	an Aesthetics Assessment (AA) network, both of which share multiple initial convolution layers. • <u>Results:</u> The results indicate that the capturing performance is benefited from aesthetic assessment. This conclusion aligns with the claims shared by pervious aesthetics-based capturing methods. Any IOT, data processing system will have the lowest threshold for latency using the image capturing method • <u>Future Scope:</u> In the proposed approach, the authors formulate aesthetics analysis as a binary classification problem (i.e., low- or high-aesthetics). However, the aesthetics assessment may be more of a ranking problem, since individuals have different aesthetics tastes but are more consistent with the relative aesthetic ranks
6.	Crop prediction using machine learning <u>Authors:</u> Madhuri Shripathi Rao, ,Arushi Singh, N.V. Subba Reddy and Dinesh U Acharya <u>DOI:</u>i:10.1088/1742-6596/2161/1/012033 (2021)	• The project aims to compare various supervised learning algorithms like KNN, Decision Tree, and Random Forest on the dataset containing 22 varieties of crops. For the Decision Tree and Random Forest Classifier, the model's performance is calculated under two criterions-Entropy and Gini Index • <u>Results:</u> In completion, we concluded that the crop prediction dataset showed the best accuracy with Random Forest Classifier both in Entropy and Gini Criterion with 99.32%. In contrast, K-Nearest Neighbor has the lowest accuracy among the three with 97.04%, and the accuracy of Decision Tree Classifier is in between KNN and Random Forest Classifier. When comparing the accuracy value, Decision Tree Gini criterion gave a better accuracy of 98.86% compared to Decision Tree Entropy Criterion • <u>Future Scope:</u> New data from the fields can be collected to get a clear image of the soil and incorporate other machine learning algorithms and deep learning algorithms such as ANN or CNN to classify more varieties of crops.
7.	Data Collection and Wireless Communication in Internet of Things (IoT) Using Economic Analysis and Pricing Models: A Survey	• This paper provides a state-of-the-art literature review on economic analysis and pricing models for data collection and wireless communication in Internet of Things (IoT). Pricing introduces another dimension of the problem that the

	Authors:N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim and Z. Han DOI:10.1109/COMST.2016.2582841 . (2016)	traditional methods of system optimization are not applicable. This is evident as a lot of researchers are studying economic issues of IoT in which many works reviewed in our survey have proposed using different approaches to address various challenges. Hence, collection of data from crops is a significant and costly challenge. • Results: This paper has provided a comprehensive survey of the economic and pricing theory as well as their applications in data collection and communication of IoT. They have presented a general architecture of an IoT system including its components and services. Important research directions for convenient and secure data collection is provided • Future Scope: There may be a situation in which the phone users are viewed s normal. but then move out of the area of interest. This may degrade the utility of the platform. Therefore, the server needs to keep track of the mobility pattern of each phone user. The lightweight triangulation method can be used to estimate the next location of each phone user. The estimation value is then considered as one of attributes to select the winners
8.	Smart Farming System: Crop Yield Prediction Using Regression Techniques Authors:Shah, A., Dubey, A., Hemnani, V., Gala, D., Kalbande, D.R. DOI:10.1007/978-981-10-8339-6_6(2018)	• This paper proposes an intelligent way to predict crop yield and suggest the optimal climatic factors to maximize crop yield. With the advancement in technology, the focus has now shifted to using machines and control systems to automate the processes and optimize productivity. The paper uses multivariate polynomial regression, support vector machine regression and random forest models to predict the crop yield per acre. The proposed method uses yield and weather data collected from United States Department of Agriculture. The various parameters included in the dataset are humidity, yield, temperature and rainfall. • Results: The predictions help the farmers choose the most suitable temperature and moisture content at which the crop yield will be optimal. The paper uses RMSE, MAE, median absolute error, and R-square values to compare between multivariate polynomial regression, support vector machine regression and random forest.

		• Future Scope: High human Involvement, high training Time, huge volume of Data and costly data feeding techniques make this approach impractical in the current market. IOT technology for data collection and feeding is priced much beyond the scope of the model prediction.
9.	Crop Prediction using Machine Learning Approaches Authors:Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R DOI:10.17577/IJERTV9IS080029(2020)	• The designed system will recommend the most suitable crop for particular land. Based on weather parameter and soil content such as Rainfall, Temperature, Humidity and pH. They are collected fromV CFarm Mandya, Government website and weather department. The system takes the required input from the farmers or sensors such as Temperature, Humidity and pH.This all inputs data applies to machinelearning predictive algorithms like Support Vector Machine (SVM) • Results:The proposed system recommends the best suitable crop for particular land by considering parameters as annual rainfall, temperature, humidity and soil pH. Among these parameters annual rainfall is predicted by system itself by using previous year data • Future Scope: Still have to collect all required data by giving GPS locations of a land and by taking access from Rain forecasting system of by the government, we can predict crops by just giving GPS location. The model can be developed to avoid over and under crisis of the food.
10.	Crop yield prediction using machine learning: A systematic literature review **Authors:** Thomas van Klompenburga , Ayalew Kassahuna , Cagatay Catalb **DOI**: 10.1016/j.compag.2020.105709 (2020)	According to the findings of this study, the selected publications employ a variety of aspects, depending on the scope of the research and the availability of data. Each research investigates yield prediction using machine learning, however the features differ. The scale, geological position, and crop of the research also varies. The features chosen are determined by the dataset's availability and the goal of the research. According to studies, models with more features do not always deliver the optimum performance for yield prediction. Models with more and fewer features should be evaluated to discover the best performing model. **Results**: After the identification of 30 papers that applied deep learning, the authors extracted and

		synthesized the applied algorithms. CNN, LSTM, and DNN algorithms are the most preferred deep learning algorithms. However, there are also other kinds of algorithms applied to this problem. **Future Scope**: The authors aim to build on the outcomes of this study and focus on the development of a DL-based crop yield prediction model.
11.	Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications **Authors:** Dhivya Elavarasan And P. M. Durairaj Vincent **DOI:** 10.1109/access.2020.2992480 (2020)	Crop yield prediction has been studied using environmental, soil, water, and crop characteristics. Deep-learning-based algorithms are widely employed in crop prediction to extract significant crop traits. To anticipate agricultural yield, the suggested work builds a Deep Recurrent Q-Network model, which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm. The reinforcement learning agent combines parametric features with the threshold to aid in crop yield prediction. **Results:** Based on the input parameters, the Q-learning network creates a crop yield prediction environment. With an accuracy of 93.7%, the suggested model outperforms existing methods in predicting crop production while keeping the original data distribution. **Future Scope:** In the future, more agricultural production prediction parameters related to pests and infestations, as well as crop damage, can be integrated in the current framework to build a more comprehensive functional model. Future improvements in the computational efficiency of the training process are also noted.
12.	A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction **Authors:** Mamunur Rashid , Bifta Sama Bari , Yusri Yusup , Mohamad Anuar Kamaruddin , And Nuzhat Khan **DOI:** 10.1109/access.2021.3075159 (2021)	Crop yield predictions are used to estimate higher crop yields using machine learning algorithms, which is one of the most difficult concerns in the agriculture sector. Because of the growing importance of agricultural yield prediction, this article presents a comprehensive analysis of the application of machine learning algorithms to forecast crop production, with a focus on palm oil yield prediction. The application of machine learning in the palm oil business, as well as a comparative study of related studies, are provided. As a result, a thorough examination of the benefits and drawbacks of machine learning-based crop production prediction, as well as correct identification of current and future agricultural industry concerns, is presented. **Results:** a prospective architecture of machine

		learning-based palm oil yield prediction has been proposed based on the critical evaluation of existing related studies. **Future Scope:** More research with a big number of features and a diverse set of prediction algorithms should be conducted.
13.	Crop Yield Prediction Using Machine Learning Algorithms **Authors:** Aruvansh Nigam , Saksham Garg ,Archit Agrawal ,Parul **DOI:** 10.1109/iciip47207.2019.89859 (2019)	This research focuses on estimating crop yield using various machine learning algorithms. The results of these strategies are compared using mean absolute error. Machine learning algorithms' predictions will assist farmers in deciding which crop to cultivate to maximise production by taking into account elements such as temperature, rainfall, acreage, and so on. **Results:** Experiments using Indian government datasets revealed that the Random Forest Regressor offers the highest yield forecast accuracy. When all parameters are pooled, the results show that Random Forest is the best classifier. **Future Scope:** Further studies with a large number of features and a wide range of prediction algorithms should be investigated.
14.	Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector **Authors:** Dr. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi , K. Neha, V.G.R.R. Devi **DOI:** 10.1109/icces48766.2020.91378 (2020)	The proposed method assists farmers in gaining an understanding of crop demand and price. It assists farmers in deciding which crop to produce in the field. The greater the increase in accuracy, the greater the profit to crop output. This work is used to obtain knowledge about the crop that can be used to make an efficient and useful harvesting. Maximum crop types have been covered by this technique. **Results:** Using data mining techniques crop yield is predicted. Here, using Random Forest algorithm for predicting the best crop yield as output. **Future Scope:** This research could be taken to the next level by developing a recommender system for agricultural production and distribution for farmers. By which farmers can make their own decisions about when and what crops to plant in order to maximise profits.
15.	Crop Yield Analysis using SIF and Climate Variables: A Case Study in Punjab, India Authors:Gautam, Pankaj Kumar Bhattacharjee, Shrutilipi DOI:10.1109/R10-	• This study investigates the correlation between the monthlyrice yield, SIF and other environmental parameters (temperature, rainfall) for Punjab district, India. The study concludesthat the considered parameters show a good correlation withthe monthly yield of rice crops.

	HTC54060.2022.9929744	• Pros:For theregression analysis, the regularized regression models (lasso,ridge, and elastic net) performs comparatively better thanthe linear regression with respect to both R2and RMSE. • Cons: This study is conducted district-wise. The results may differ when implemented on a larger national scale.
16.	Integrating Crop Simulation and Machine Learning Models to Improve Crop Yield Prediction Authors:Saiara Samira Sajid, Isaiah Huber, SotiriosArchontoulis,Guiping Hu DOI:10.1109/SOSE55472.2022.9812678 (2022)	• This research focuses on identifying a robust crop prediction model that can be applied on a large scale. To build the prediction model results from crop simulation software (APSIM) were combined with ML models. Finally, an optimized weighted ensemble model is developed. • Pros:In a previous study, optimum weights were selected by minimizing MSE, which had limitations in calculating optimum weights when applied to a large scale. However, considering minimizing RMSE as an objective function overcomes the concern and can be applied on a wide scale. • Cons:The correlation of model performance with soil and weather properties was evaluated, and it was found that correlations are random. Thus, these attributes which influence crop yield in reality, have no effect in the ML model.
17.	A Review of Crop Yield Prediction Strategies based on Machine Learning and Deep Learning Authors:R S Renju, P S Deepthi, M T Chitra DOI:10.1109/IC3SIS54991.2022.9885325 (2022)	• This paper presented an overview of various crop yield forecasting models in the recent literature. The study reveals that each crop yield prediction employed more elements based on the scope of the work and the availability of data. • Even though the data features vary, each work presented in this paper concentrates on machine learning and deep learning techniques for obtaining accurate crop yield prediction. The major feature selection concerns the availability of relevant datasets and attaining the research goal.
18.	Crop Yield Prediction using Random Forest Algorithm Authors:Rajesh Yamparla,Harisa Sultana Shaik, Naga Sai PravallikaGuntaka, Pallavi Marri,SrilakshmiNallamothu	• This paper proposed a model that predicts crop yield using Random Forest Algorithm. The above study shows us that the Random ForestAlgorithm when comparing to the above models theRandom Forest has the Highest Accuracy of 95%.

	DOI:10.1109/ICCES54183.2022.9835756 (2022)	• Pros:Therandom forest method offers a high level of accuracysince it uses a bagging strategy. This strategy reducesmodel complexity, resulting in overfitting of data fortraining • Cons: As the scale of applicability increases, the prediction accuracy becomes inferior to gradient-boosted trees. Thus, realising the scale of model implementation is important.
19.	Deep-LSTM Model for Wheat Crop Yield Prediction in India Authors:Preeti Saini; Bharti Nagpal DOI:10.1109/CCiCT56684.2022.00025 (2022)	• This Study highlights the efficiency of the DeepLSTM architecture for the Wheat Yield Prediction in India using a Historical dataset. The proposed model outperforms the existing GPR and Holt-Winter methods with a lower RMSE Value of 0.20, 0.51 & 0.61 respectively. • Pros: . The advantage of deep neural networks is their ability to find non-linear patterns in real datasets using predictive modelling. • Cons: In order to obtain the accuracy achieved by the authors, the pre-requisites for developing the model outweigh the goal, thus making such a model impractical.
20.	Early Prediction of Crop Yield in India using Machine Learning. Authors: Ankita Sharma, AnushthaTamrakar, SourajitaDewasi, Nenavath Srinivas Naik DOI:10.1109/TENSYMP54529.2022.9864490 (2022)	• The proposed model in this research considers the data related to the district, season, year, production area, and crop type, and then it predicts crop yield. The result shows that accurate prediction can be attained using RFR. • Pros: This model will also be able to recommend crops given set of environmental conditions. Also, the inclusion of attributes of weather, geo-climatic situations and soil minerals in the classification model gave the farmers the best crop to grow with a set of characteristics of surroundings with the best results. • Cons: The variance present in the model is not exactly low, and bordering on the line between impractical and practicality.

4. Proposed Work:

The Proposed system will predict the crop yield for a particular crop based on weather parameters such as Temperature and Rainfall and also the amount of Pesticides.

Technology Used:

- Machine Learning
- Ensemble supervised Learning
- Extra Trees Regressor

Tools and Packages used:

1. **Python**

Python is an object-oriented, high-level programming language with dynamic semantics that is interpreted. Python's concise, easy-to-learn syntax prioritises readability, lowering software maintenance costs. Python has support for modules and packages, which promotes programme modularity and code reuse.

The Python libraries we utilised in our project are displayed below.

A. Pandas:

pandas is a software library present in the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

B. NumPy:

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

C. Matplotlib:

Matplotlib is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc. It supports a very wide variety of graphs and plots namely - histogram, bar charts, power spectra, error charts etc. It is used along with NumPy.

D. Seaborn:

Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with dataframes and the Pandas library. The graphs created can also be customized easily.

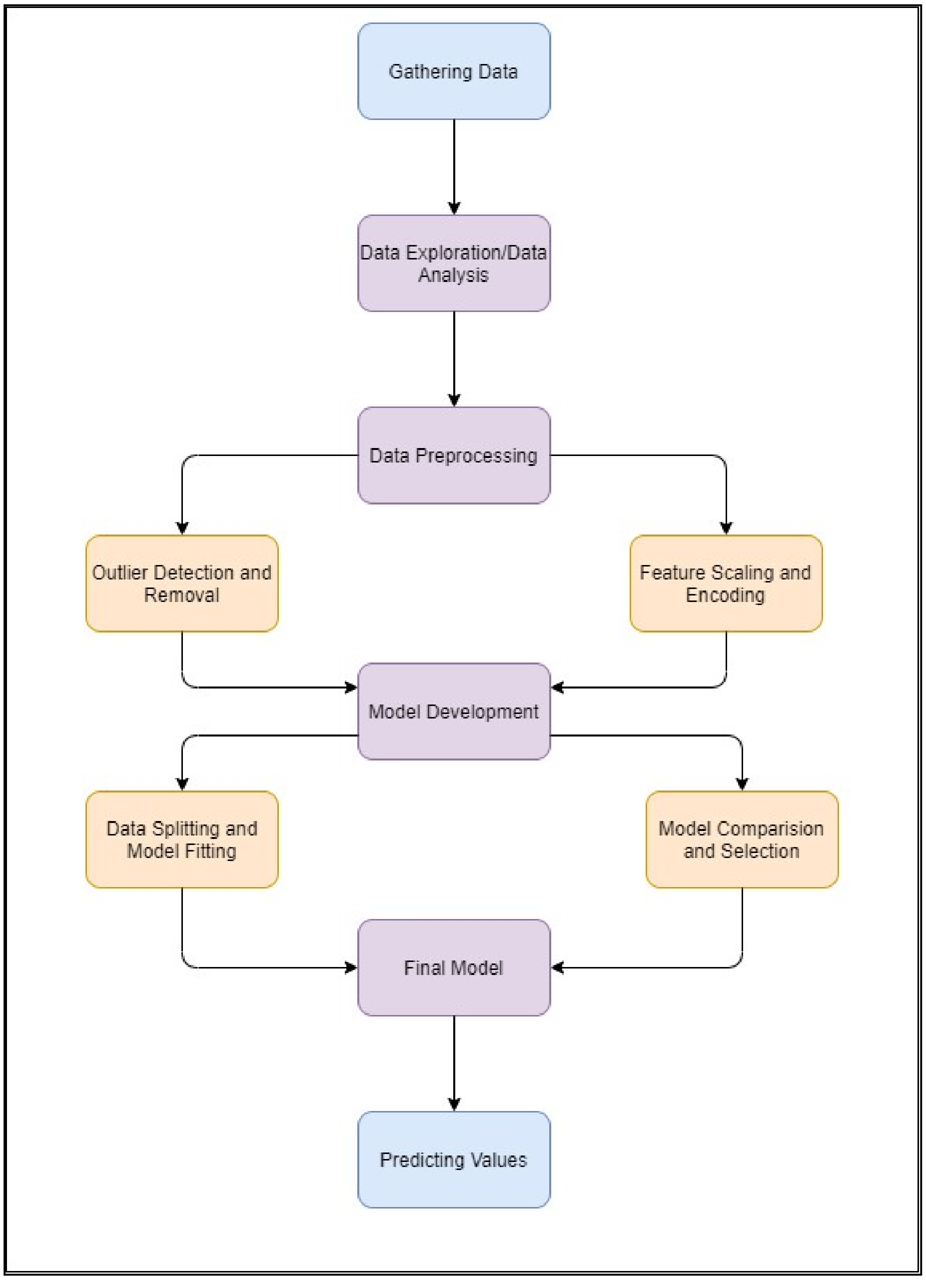
E. Scikit-learn (Sklearn):

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

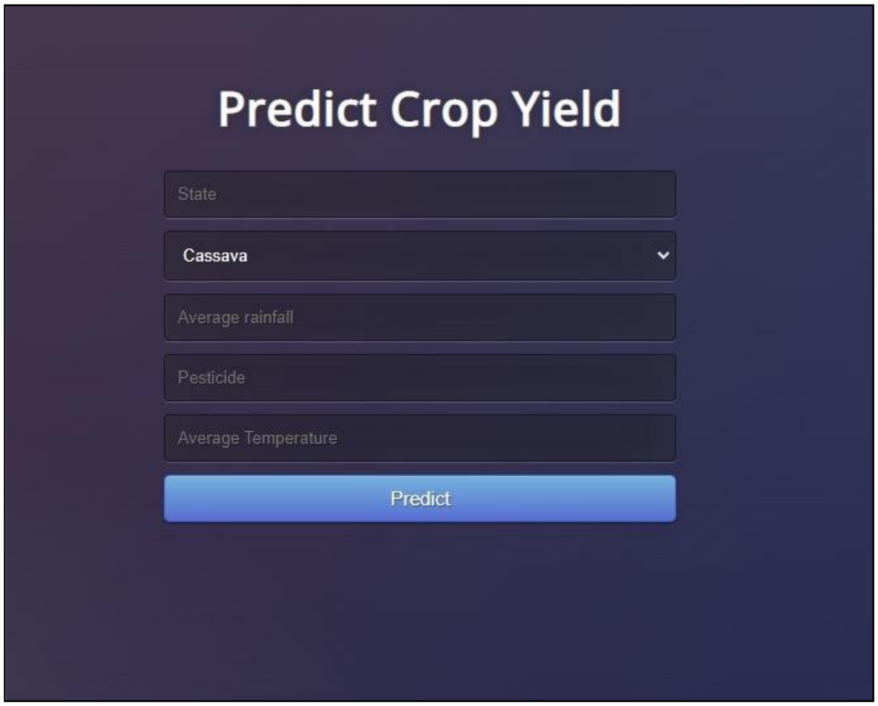
2. Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation

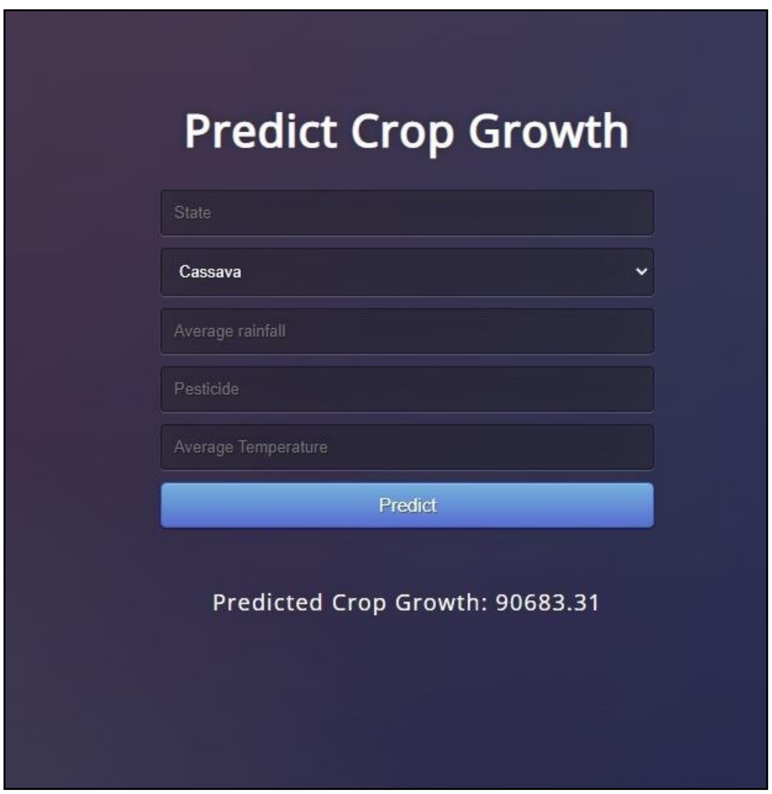
Model Design



4.4 User Interface



After Prediction:



Algorithm: Extra Trees Regression

Extra Trees, or Extremely Randomized Trees, is an ensemble machine learning algorithm. It is a decision tree ensemble and is connected to other decision tree ensemble methods such as bootstrap aggregation (bagging) and random forest. Extra Trees generates a huge number of unpruned decision trees from the training dataset. In the case of regression, predictions are formed by averaging the forecast of the decision trees, and in the case of classification, predictions are made by employing majority voting.

- o Regression: Forecasts made by averaging decision tree predictions.
- o Classification: Predictions from decision trees made by majority vote.

Unlike bagging and random forest, which create each decision tree using a bootstrap sample of the training dataset, the Extra Trees approach fits each decision tree on the entire training dataset. The Extra Trees algorithm, like random forest, will randomly sample the characteristics at each split point of a decision tree. The Extra Trees algorithm chooses a split point at random, as opposed to random forest, which utilises a greedy algorithm to choose an ideal split point.

The Extra-Trees approach use the traditional top-down procedure to construct an ensemble of unpruned decision or regression trees. Its two primary differences from other tree-based ensemble approaches are that it separates nodes by selecting cut-points completely at random and that it grows the trees using the entire learning sample (rather than a bootstrap replica).

As a result, the technique has three key hyperparameters to tune: the number of decision trees in the ensemble, the number of input features to randomly select and consider for each split point, and the minimum amount of samples required in a node to establish a new split point. It contains two parameters: K, which is the number of randomly picked characteristics at each node, and nmin, which is the minimum sample size for splitting a node. The number of trees in this ensemble is denoted by M. The random selection of split points makes the decision trees in the ensemble less correlated, but it raises the algorithm's variance. This increase in variance can be compensated for by increasing the number of trees in the ensemble. The parameters K, nmin, and M have diverse effects: K controls the strength of the attribute selection process, nmin determines the intensity of averaging output noise, and M affects the strength of the ensemble model aggregation's variance reduction.

Why use Extra Trees Regression ?

Like the random forests technique, the extra trees algorithm generates a large number of decision trees, but the sampling for each tree is random and without replacement. This generates a dataset with unique samples for each tree. For each tree, a particular number of features are chosen at random from the complete collection of features. The random selection of a splitting value for a feature is the most essential and distinguishing aspect of extra trees. Instead of computing a locally optimal value for splitting the data using Gini or entropy, the algorithm chooses a split value at random. As a result, the trees are diverse and unrelated.

Algorithm	Accuracy (Training Set)	Accuracy (Test Set)	Average Accuracy
Linear Reg.	0.71974	0.72758	0.72366
Random Forest Reg.	0.99914	0.98990	0.99452
Bayesian Ridge Reg.	0.7197	0.72760	0.72865
Decision Tree Reg.	1.0	0.98278	0.99139
Gradient Boost Reg.	0.79935	0.816	0.80767
AdaBoost Reg.	0.56572	0.58611	0.57591
Extra Trees Reg.	1.0	0.99349	0.99674

By comparing the accuracy scores of each model, we can say that Extra Trees Regression Algorithm works best with our data. Thus we will be implementing the above mentioned algorithm in our model to predict crop yield.

Algorithm	R2 Scores
Linear Reg.	0.6784
Random Forest Reg.	0.9839
Bayesian Ridge Reg.	0.6784
Decision Tree Reg.	0.9802
Gradient Boost Reg.	0.9036
AdaBoost Reg.	0.6611
Extra Trees Reg.	0.9847

R2 is a statistical measure ranging from 0 to 1 that calculates how close a regression line is to the data to which it is fitted. If it's a 1, the model predicts 100% of the variance in the data; if it's a 0, the model predicts none of the variance. Extra Trees Regressor gets the greatest R2 score of 98%, with Random Forest Regressor coming in second.

When the outcomes of the Random Forest and Extra Tree algorithms are compared, they are nearly identical. It is worth observing the difference in execution time, where Extra Trees is significantly faster. As a result, while deciding which of the two ensembles to employ, it appears a good idea to utilise Extra Trees because the same result is reached faster.

5. Experiments/Results:
Gathering Data

The dataset includes the following crops:

- Cassava
- Maize
- Plantains and others
- Potatoes
- Rice, paddy
- Sorghum
- Soybeans
- Sweet potatoes
- Wheat
- Yams

Dataset Parameters

	Area	Item	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp
0	Andaman & Nicobar Islands	Cassava	1990	31400	1996	12004.33	19.64
1	Andaman & Nicobar Islands	Maize	1990	20052	1996	12004.33	19.64
2	Andaman & Nicobar Islands	Plantains and others	1990	394286	1996	12004.33	19.64
3	Andaman & Nicobar Islands	Potatoes	1990	210685	1996	12004.33	19.64
4	Andaman & Nicobar Islands	Rice, paddy	1990	31111	1996	12004.33	19.64

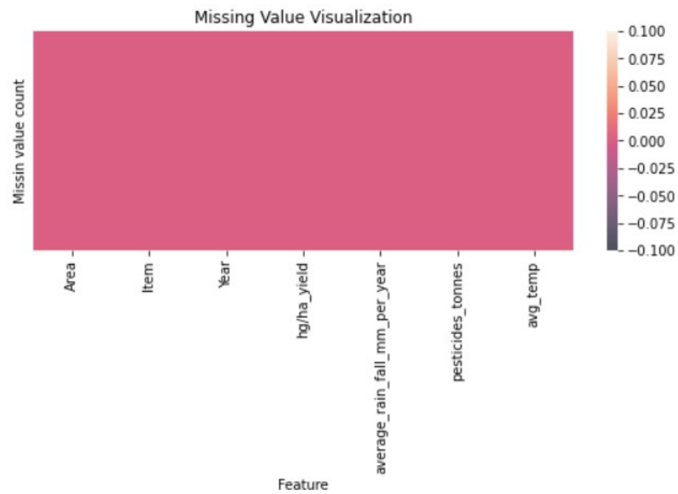
The high variance in the values for each column can be seen; later on, we can accommodate for that via scaling.

	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp
count	20847.000000	20847.000000	20847.000000	20847.000000	20847.000000
mean	2001.472970	79402.027102	1174.055835	48077.171702	20.980473
std	7.060078	86530.686797	707.913760	65751.304038	5.904574
min	1990.000000	866.000000	51.000000	0.920000	1.610000
25%	1995.000000	20728.000000	591.000000	8674.580000	17.330000
50%	2001.000000	38260.000000	1083.000000	34468.930000	22.120000
75%	2008.000000	107692.000000	1738.000000	58349.440000	25.960000
max	2013.000000	457565.000000	3240.000000	367778.000000	29.410000

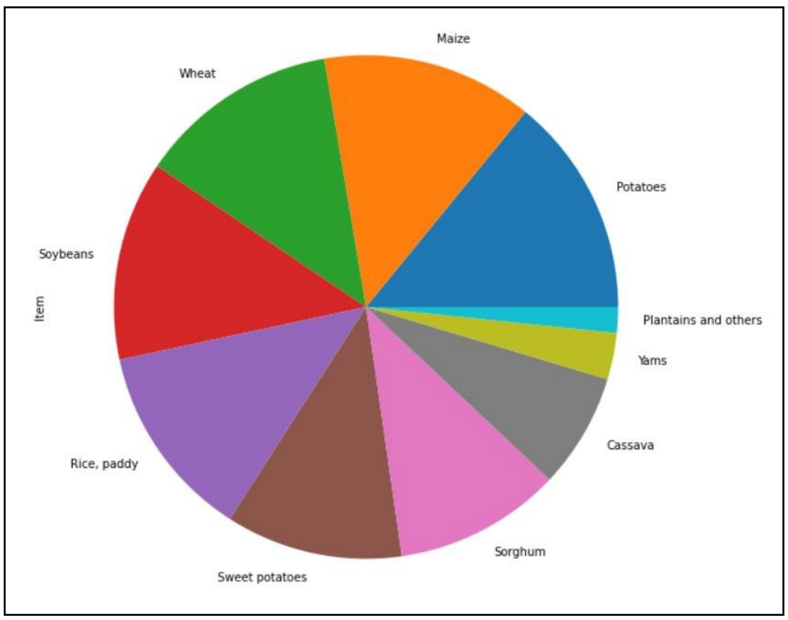
Data Exploration

<u>Checking For Missing Values: There are no missing values as seen below</u>

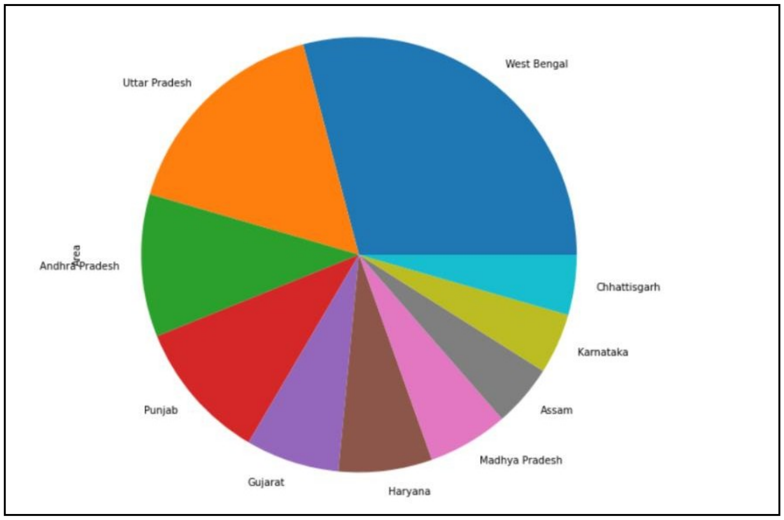
```
Area                             0
Item                             0
Year                             0
hg/ha_yield                      0
average_rain_fall_mm_per_year    0
pesticides_tonnes                0
avg_temp                         0
dtype: int64
```



<u>The yield per crop:</u>



The amount of yield per area:



The Numerical Features and Categorical Features:

```
Numerical Features:  ['Year', 'hg/ha_yield', 'average_rain_fall_mm_per_year', 'pesticides_tonnes', 'avg_temp']
Categorical Features:  ['Area', 'Item']
```

The crops:

Item	Area	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp
Cassava	1541	1541	1541	1541	1541	1541
Maize	2825	2825	2825	2825	2825	2825
Plantains and others	345	345	345	345	345	345
Potatoes	2940	2940	2940	2940	2940	2940
Rice, paddy	2618	2618	2618	2618	2618	2618
Sorghum	2231	2231	2231	2231	2231	2231
Soybeans	2673	2673	2673	2673	2673	2673
Sweet potatoes	2366	2366	2366	2366	2366	2366
Wheat	2687	2687	2687	2687	2687	2687
Yams	621	621	621	621	621	621

The top 10 States according to Highest Yield:

```
Area
West Bengal        327420324
Uttar Pradesh      167550306
Andhra Pradesh     130788528
Haryana            124470912
Gujarat            109111062
Punjab              73897434
Madhya Pradesh      69193506
Jammu & Kashmir     55419990
Karnataka           52263950
Tamil Nadu          46773540
Name: hg/ha_yield, dtype: int64
```

The Top States according to yield of a particular crop:

```
Item            Area
Cassava         West Bengal        142810624
Potatoes        West Bengal         92122514
                Uttar Pradesh       49602168
                Jammu & Kashmir     46705145
                Gujarat             45670386
Sweet potatoes  West Bengal         44439538
Potatoes        Haryana             42918726
                Andhra Pradesh      42053880
Sweet potatoes  Andhra Pradesh      35808592
                Gujarat             35550294
Name: hg/ha_yield, dtype: int64
```

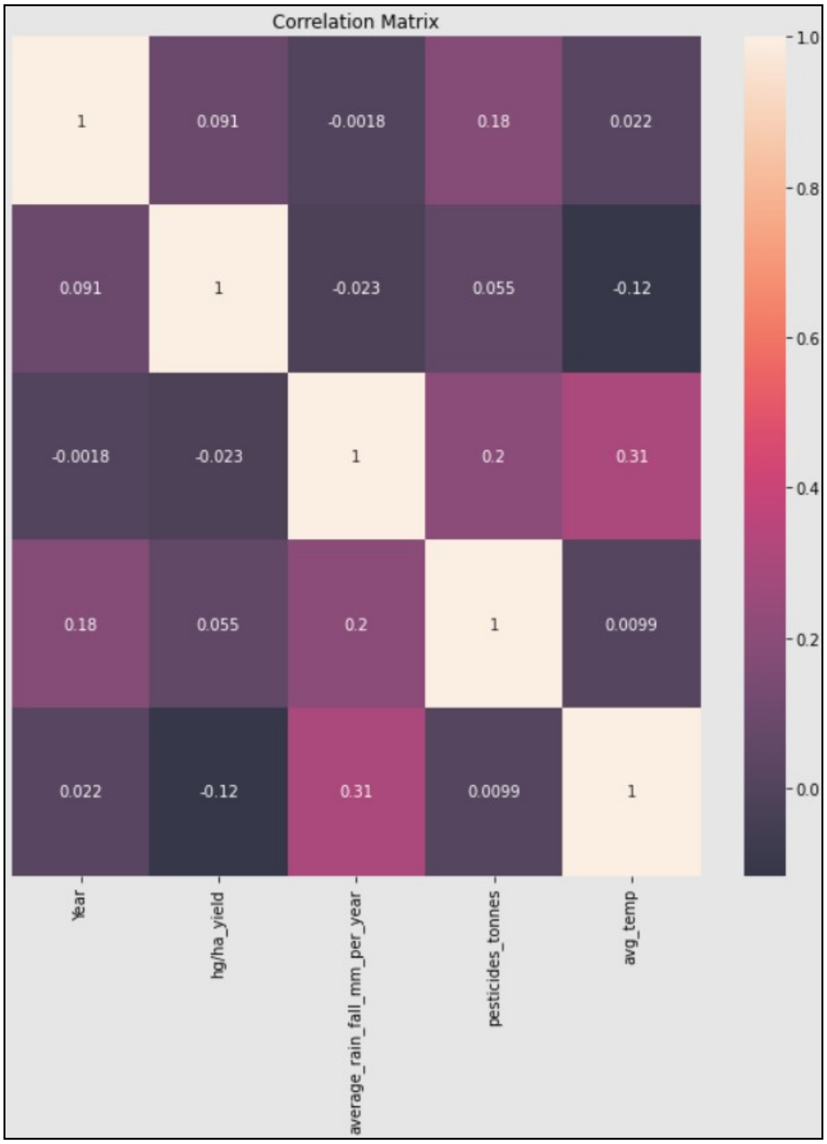
Cassava and potato production in West Bengal is the highest. Potatoes appear to be the most prevalent crop in the sample, accounting for the highest share in four countries.

The final dataframe spans the years 1990 to 2013, totaling 23 years of data for 29 States and Union Terrritories

Exploring the relationships between the dataframe's columns, an useful technique to rapidly verify correlations between columns is to visualise the correlation matrix as a heatmap.

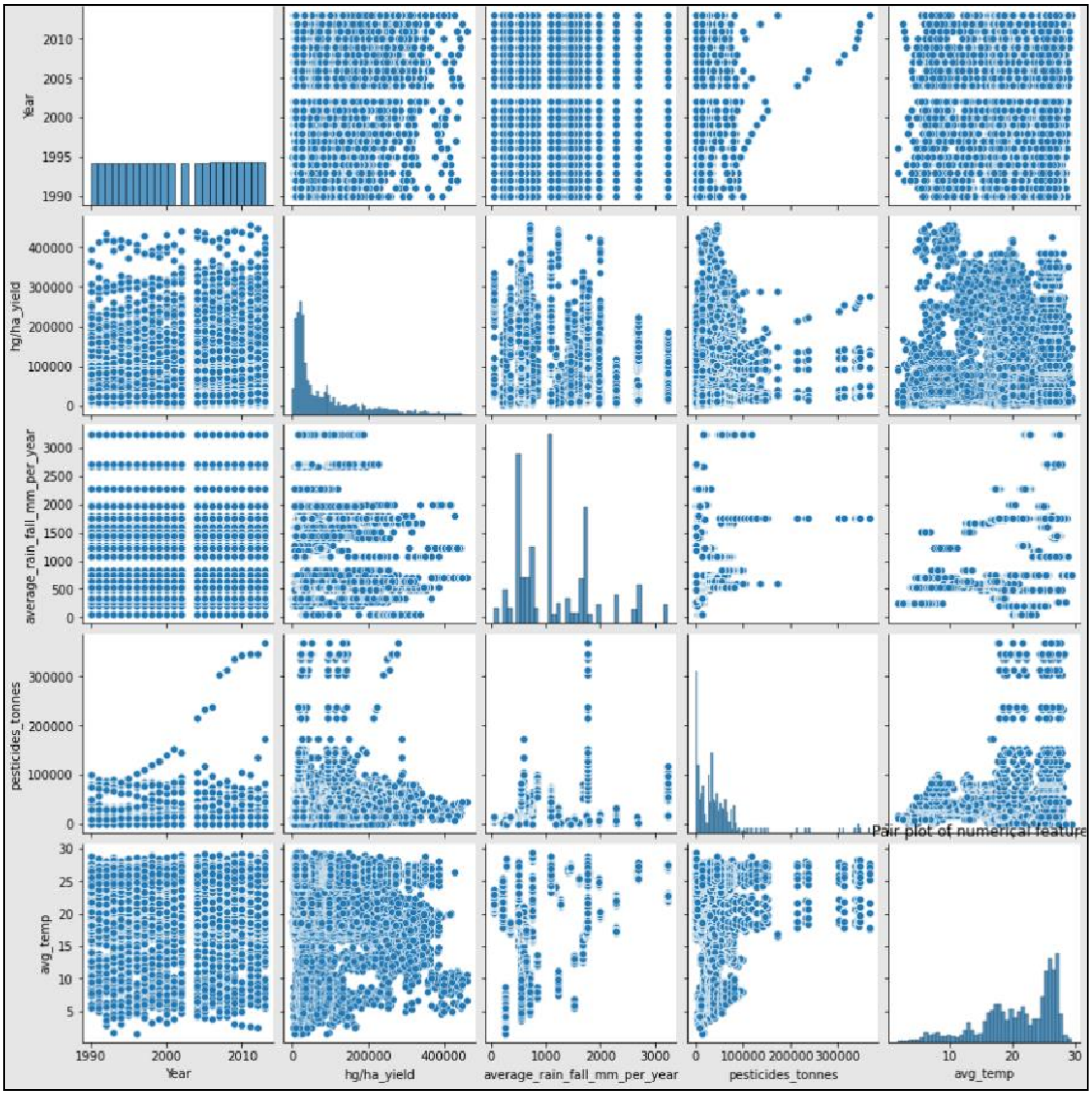
Correlation Matrix

	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp
Year	1.000000	0.091020	-0.001776	0.177390	0.021680
hg/ha_yield	0.091020	1.000000	-0.023346	0.054912	-0.116754
average_rain_fall_mm_per_year	-0.001776	-0.023346	1.000000	0.199511	0.306112
pesticides_tonnes	0.177390	0.054912	0.199511	1.000000	0.009919
avg_temp	0.021680	-0.116754	0.306112	0.009919	1.000000

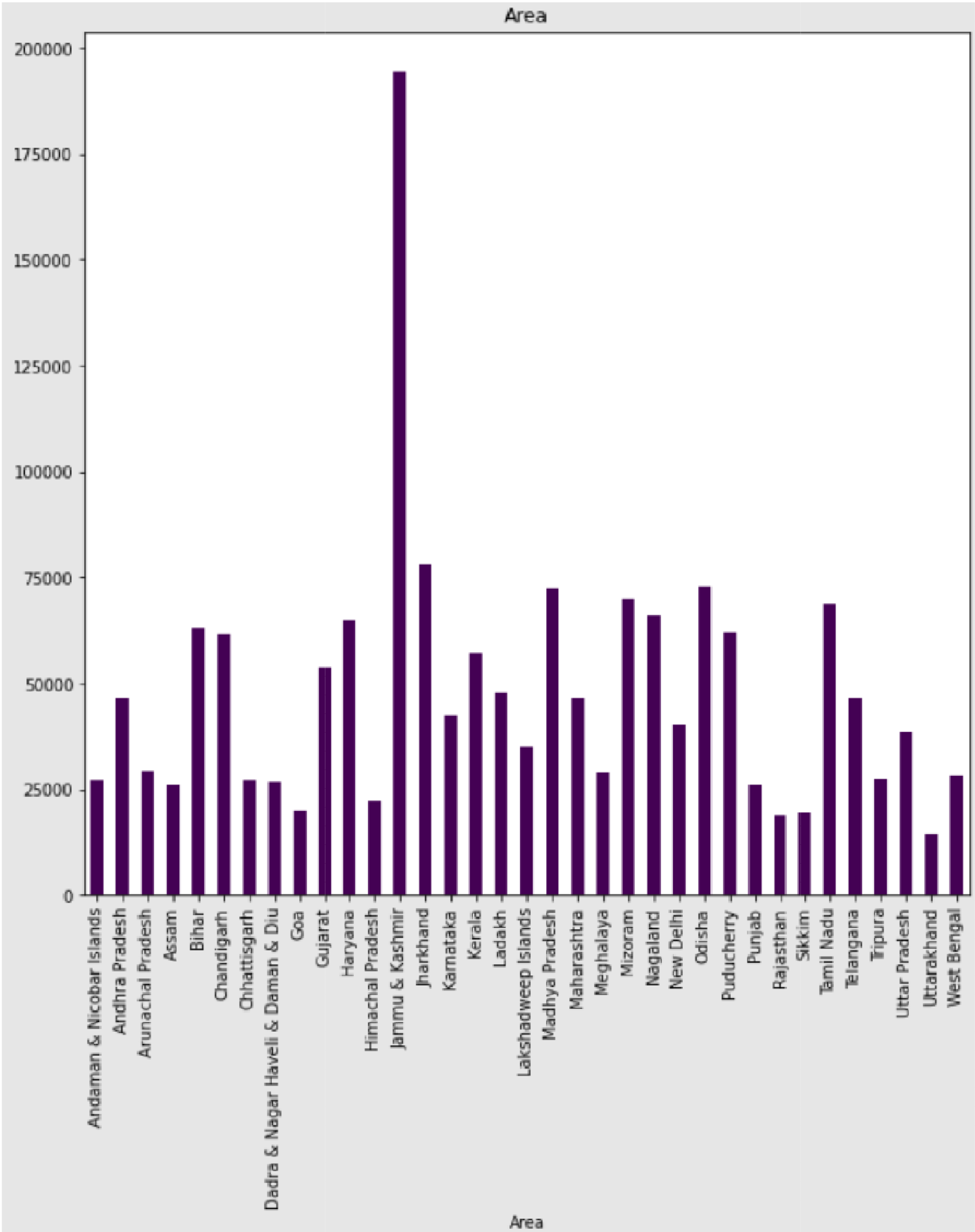


The above correlation map shows that there is no association between any of the columns in the dataframe.

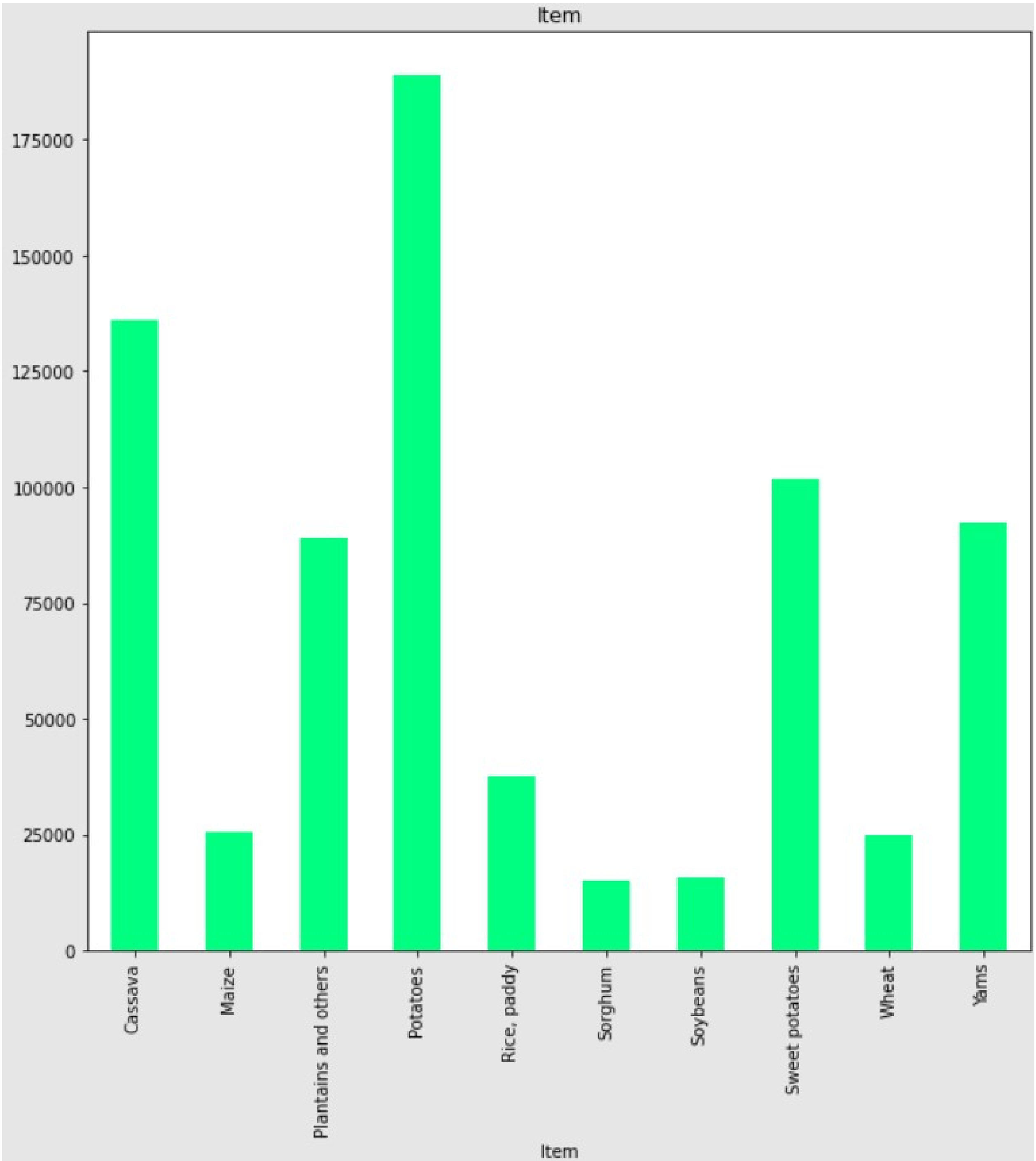
Pair Plot of Numerical Features



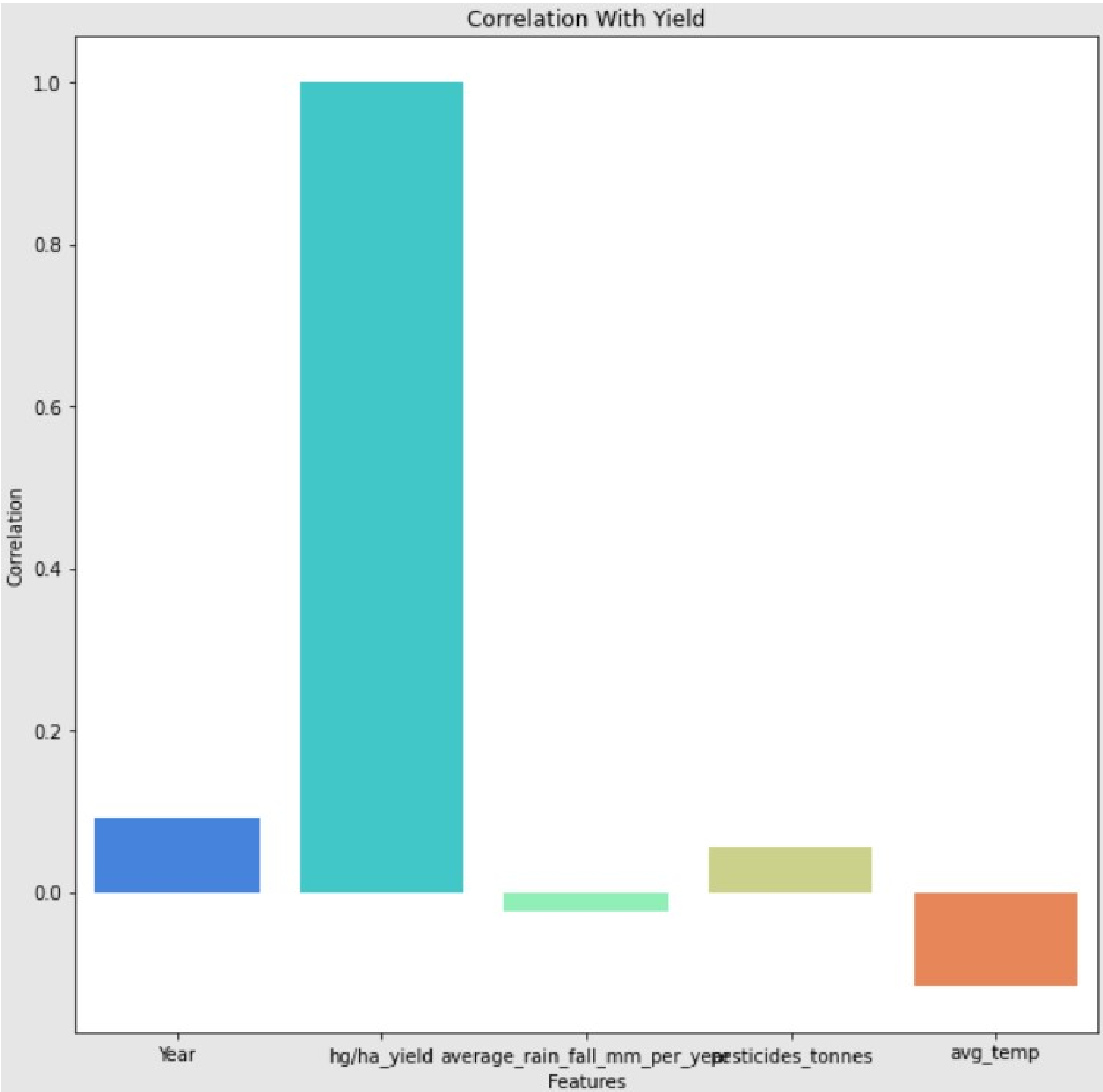
Area of Each State



Yield Per Crop



Correlation With Yield

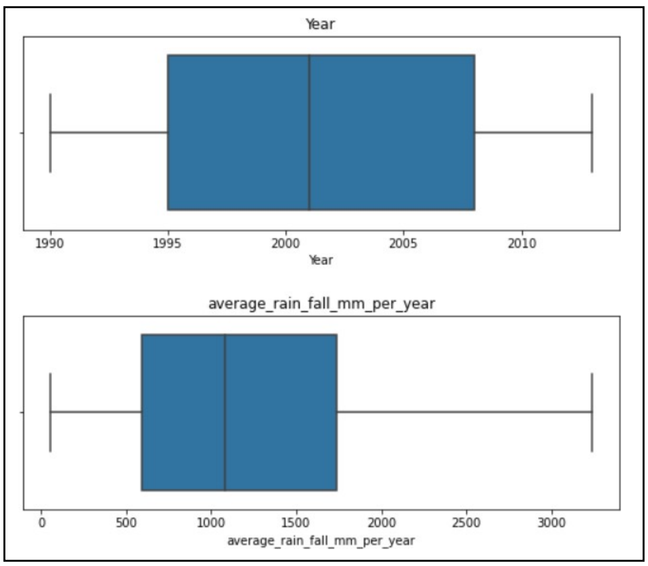


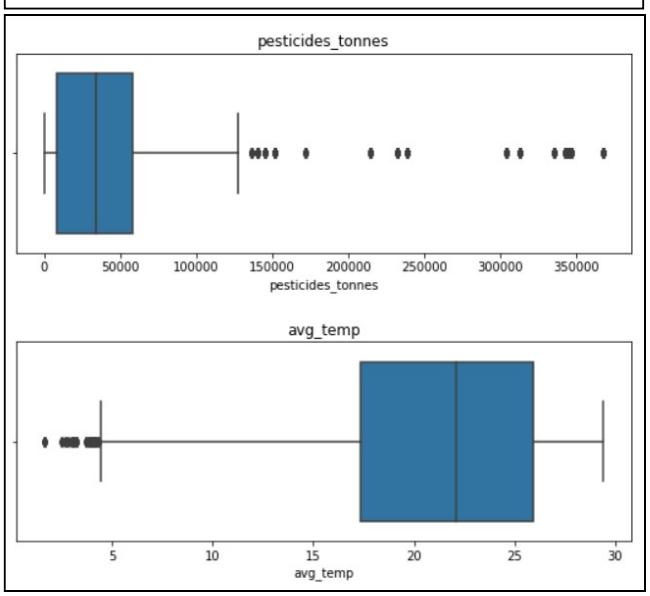
Data Preprocessing

Data preprocessing is a technique for converting raw data into a clean data set. In other words, anytime data is acquired from various sources, it is obtained in raw format, which makes analysis impossible.

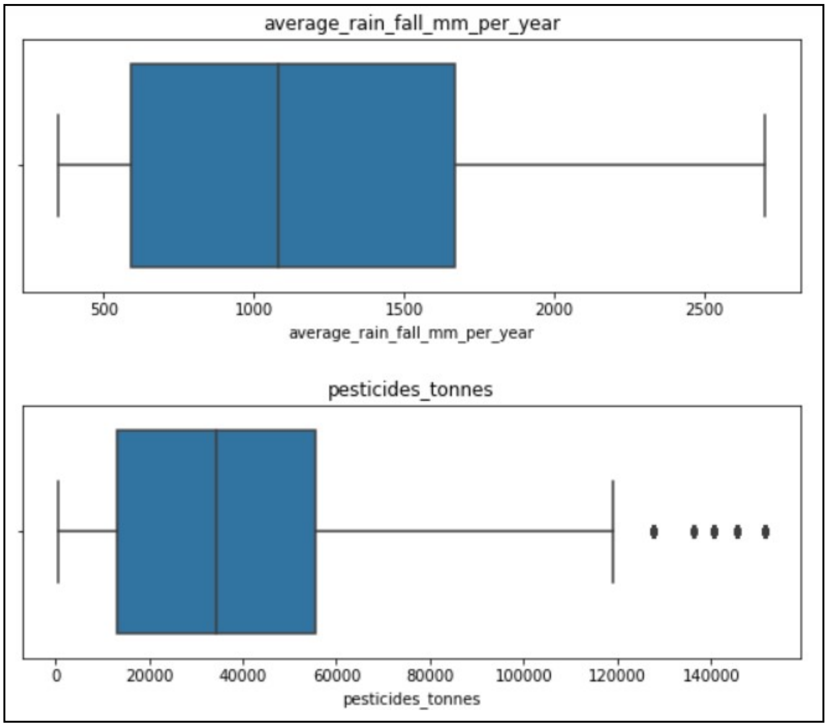
Outlier Detection and Removal

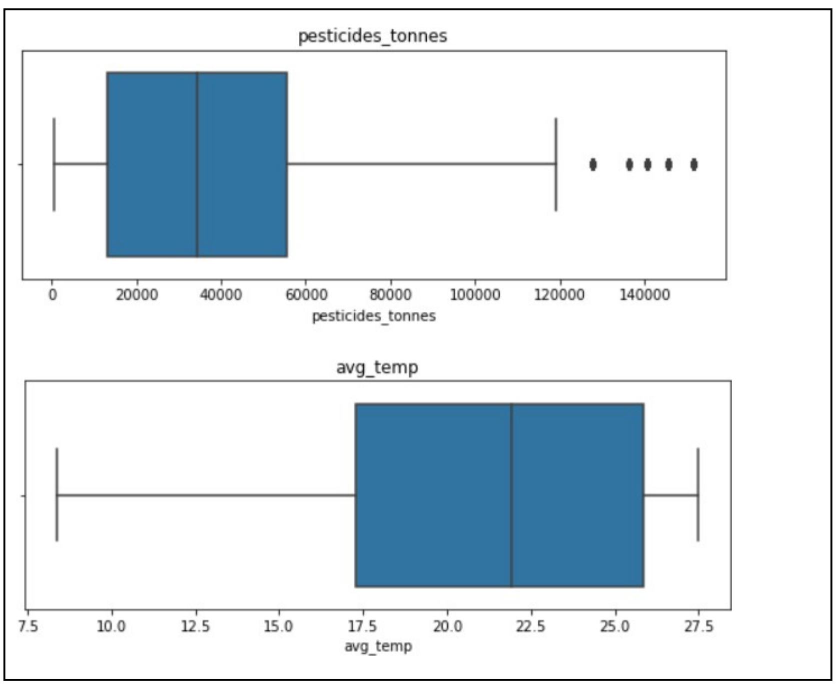
Detection: Detecting the outliers





Removal: Most of the outliers have been deducted





Feature Scaling and Encoding

Encoding

The dataframe contains two category columns; categorical data are variables that carry label values rather than numeric values. The number of available values is frequently limited to a specific set, such as the values for items and nations in this case. Many machine learning algorithms cannot operate directly on label data. They require that all input and output variables be numeric.

This means that category data must be transformed into numerical data. One hot encoding is a procedure that converts categorical variables into a form that can be fed into ML algorithms to help them predict better. One-Hot Encoding will be utilised to transform these two columns to a one-hot numeric array for this purpose.

The numerical value of the entry in the dataset is represented by the categorical value. This encoding generates a binary column for each category and returns the results as a matrix.

```
array([[1.0, 0.0, 0.0, ..., 1996, 12004.33, 19.64],
       [0.0, 1.0, 0.0, ..., 1996, 12004.33, 19.64],
       [0.0, 0.0, 1.0, ..., 1996, 12004.33, 19.64],
       ...,
       [0.0, 0.0, 0.0, ..., 1083, 45620.0, 27.44],
       [0.0, 0.0, 0.0, ..., 1083, 45620.0, 27.44],
       [0.0, 0.0, 0.0, ..., 1083, 45620.0, 26.99]], dtype=object)
```

```
array([[1.0, 0.0, 0.0, ..., 1996, 12004.33, 19.64],
       [0.0, 1.0, 0.0, ..., 1996, 12004.33, 19.64],
       [0.0, 0.0, 1.0, ..., 1996, 12004.33, 19.64],
       ...,
       [0.0, 0.0, 0.0, ..., 1083, 45620.0, 27.44],
       [0.0, 0.0, 0.0, ..., 1083, 45620.0, 27.44],
       [0.0, 0.0, 0.0, ..., 1083, 45620.0, 26.99]], dtype=object)
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	hg/ha_yield
0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1996	12004.33	19.64	31400
1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1996	12004.33	19.64	20052
2	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1996	12004.33	19.64	394286
3	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1996	12004.33	19.64	210685
4	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0	1996	12004.33	19.64	31111

Scaling

Looking at the dataset above, we can see that the magnitudes, units, and ranges of the features vary greatly. High magnitude features will weigh far more heavily in distance calculations than low magnitude features. To counteract this effect, we must bring all features to the same magnitude level. This can be accomplished through scaling.

	Area	Item	Year	hg/ha_yield	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp
0	Andaman & Nicobar Islands	Cassava	1990	31400	3.501698	0.382409	597
1	Andaman & Nicobar Islands	Maize	1990	20052	3.501698	0.382409	597
2	Andaman & Nicobar Islands	Plantains and others	1990	394286	3.501698	0.382409	597
3	Andaman & Nicobar Islands	Potatoes	1990	210685	3.501698	0.382409	597
4	Andaman & Nicobar Islands	Rice, paddy	1990	31111	3.501698	0.382409	597

After removing the year column and scaling all values in features, the resulting array will look like this:

```
array([[1.        , 0.        , 0.        , ..., 0.60990906, 0.03263773,
        0.64856115],
       [0.        , 1.        , 0.        , ..., 0.60990906, 0.03263773,
        0.64856115],
       [0.        , 0.        , 1.        , ..., 0.60990906, 0.03263773,
        0.64856115],
       ...,
       [0.        , 0.        , 0.        , ..., 0.32361242, 0.12404003,
        0.92913669],
       [0.        , 0.        , 0.        , ..., 0.32361242, 0.12404003,
        0.92913669],
       [0.        , 0.        , 0.        , ..., 0.32361242, 0.12404003,
        0.91294964]])
```

Model Development

Data Splitting and Model Fitting

The dataset will be divided into two parts: training and testing. Because training the model requires as many data points as feasible, the data is frequently split inequitably. The most typical train/test splits are 70/30 or 80/20.

The training dataset is the first dataset used to teach the machine learning algorithm to learn and make correct predictions. (80% of the dataset is used for training.) The test dataset, on the other hand, is used to evaluate how well the ML algorithm is trained with the training dataset.

```
Test Size        3244
Train Size       12972
```

Model Comparision and Selection

R2 Scores of each Model have been Calculated and Extra Trees Regressor turned out to have the Highest R2 Scores.

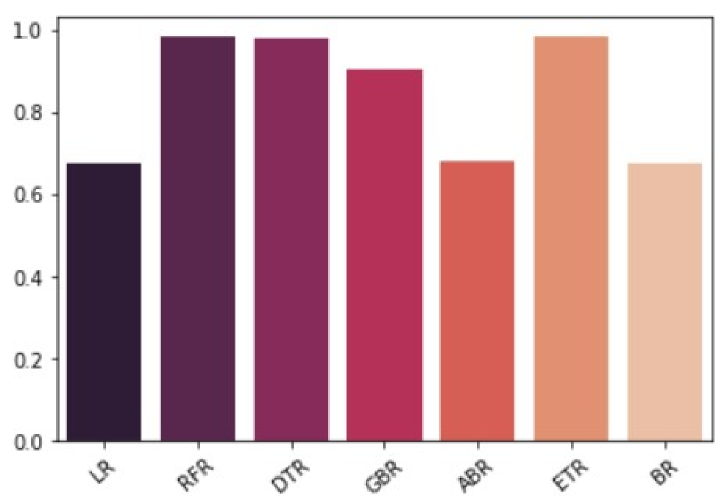
```
['LinearRegression', 0.6784745906613432]
['RandomForestRegressor', 0.9839458879759438]
['DecisionTreeRegressor', 0.9794086387128991]
['GradientBoostingRegressor', 0.9036107955720422]
['AdaBoostRegressor', 0.7092423981218265]
['ExtraTreesRegressor', 0.9847070216771822]
['BayesianRidge', 0.6784848390170964]
```

Accuracy Scores of each Model:

From The Graph and Accuracy Scores we can see that Extra trees Regressor has the best scores

```
LinearRegression()
Accuracy on training set:  0.677309344457742
Accuracy on testing set:  0.6784745906613432
RandomForestRegressor()
Accuracy on training set:  0.9974768194101045
Accuracy on testing set:  0.9840424010561224
DecisionTreeRegressor()
Accuracy on training set:  0.9996459042938979
Accuracy on testing set:  0.9797617520288467
GradientBoostingRegressor()
Accuracy on training set:  0.9113005517554312
Accuracy on testing set:  0.9036107955720422
AdaBoostRegressor()
Accuracy on training set:  0.7052421855371847
Accuracy on testing set:  0.704031108217988
ExtraTreesRegressor()
Accuracy on training set:  0.9996459042486353
Accuracy on testing set:  0.9847358318870867
BayesianRidge()
Accuracy on training set:  0.6773092411334823
Accuracy on testing set:  0.6784848390170964
```

Graph of Accuracy Scores:

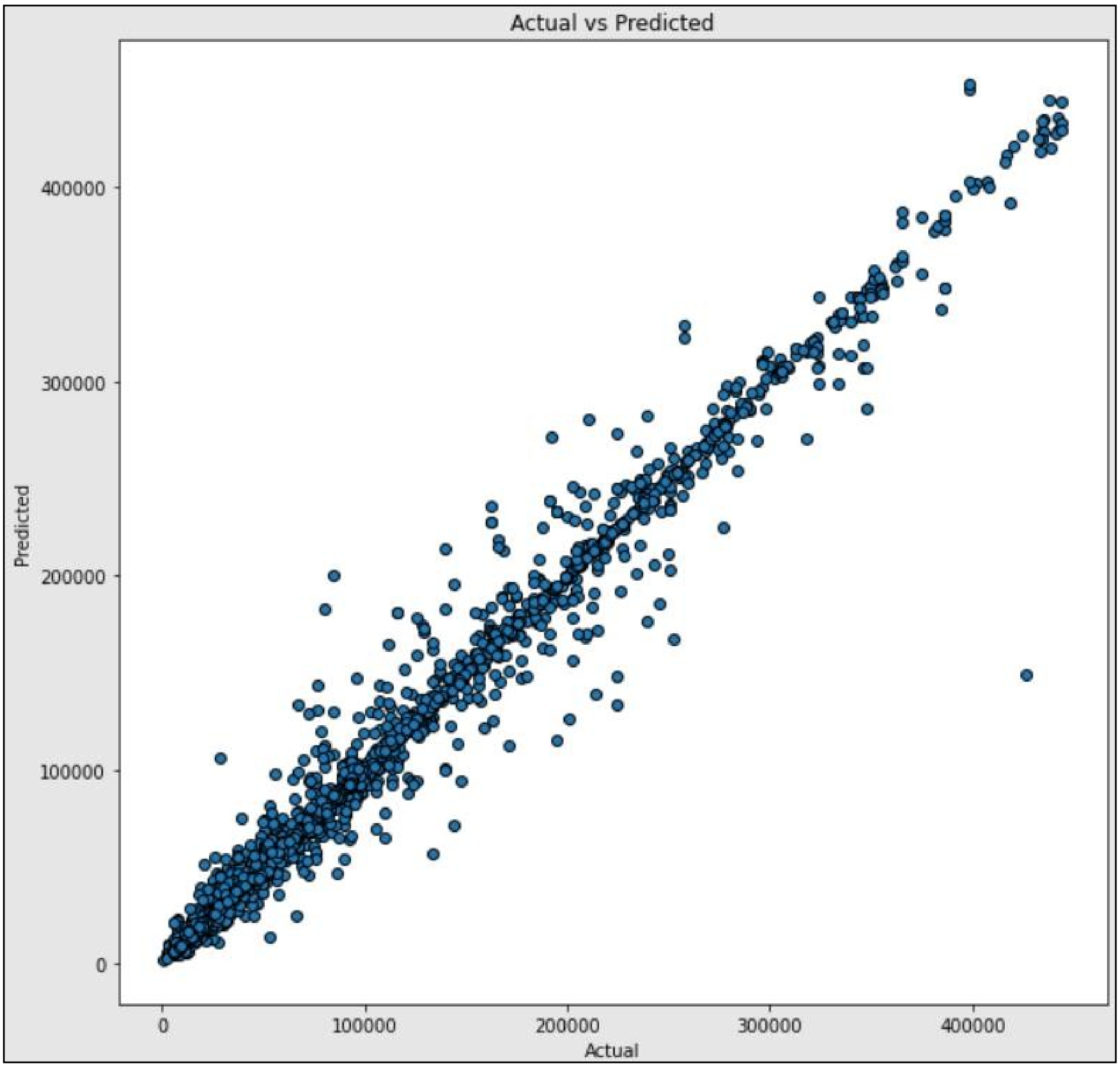


Final Model

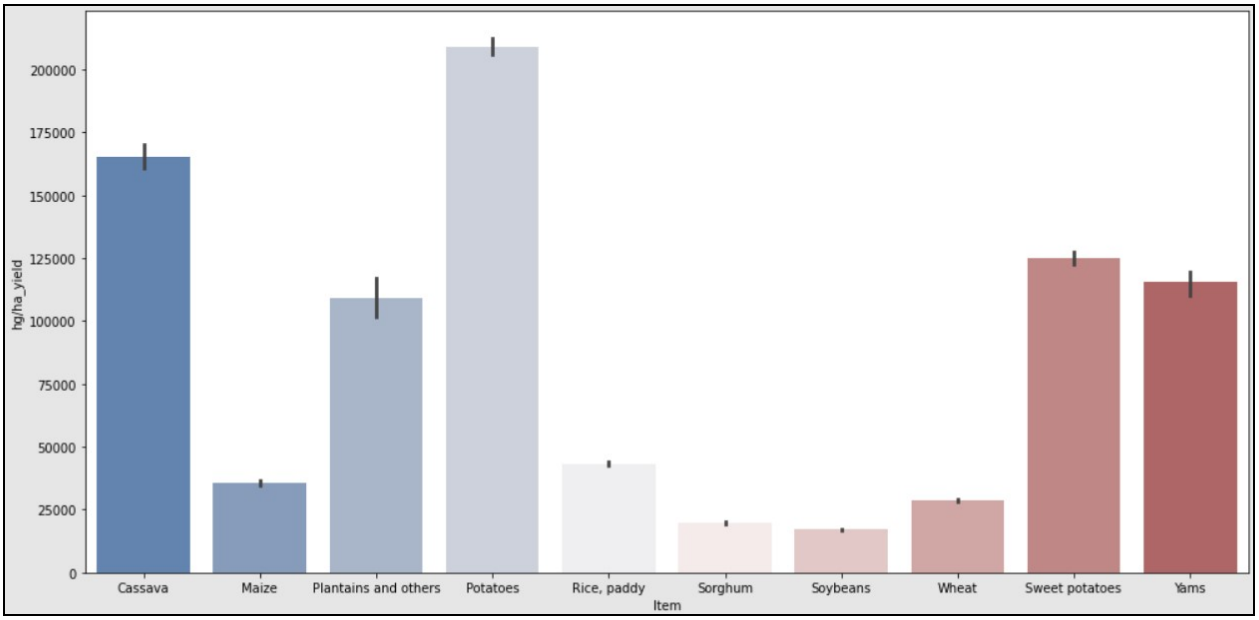
Calculating MSE ,MAE and Accuracy Scores

```
Mean Squared Error:  112389115.34444849
Mean Absolute Error:  3708.775302957634
Accuracy Score:  0.9847752004804826
```

Scatter Plot of the model's actual values against the predicted ones



Yield of Crops

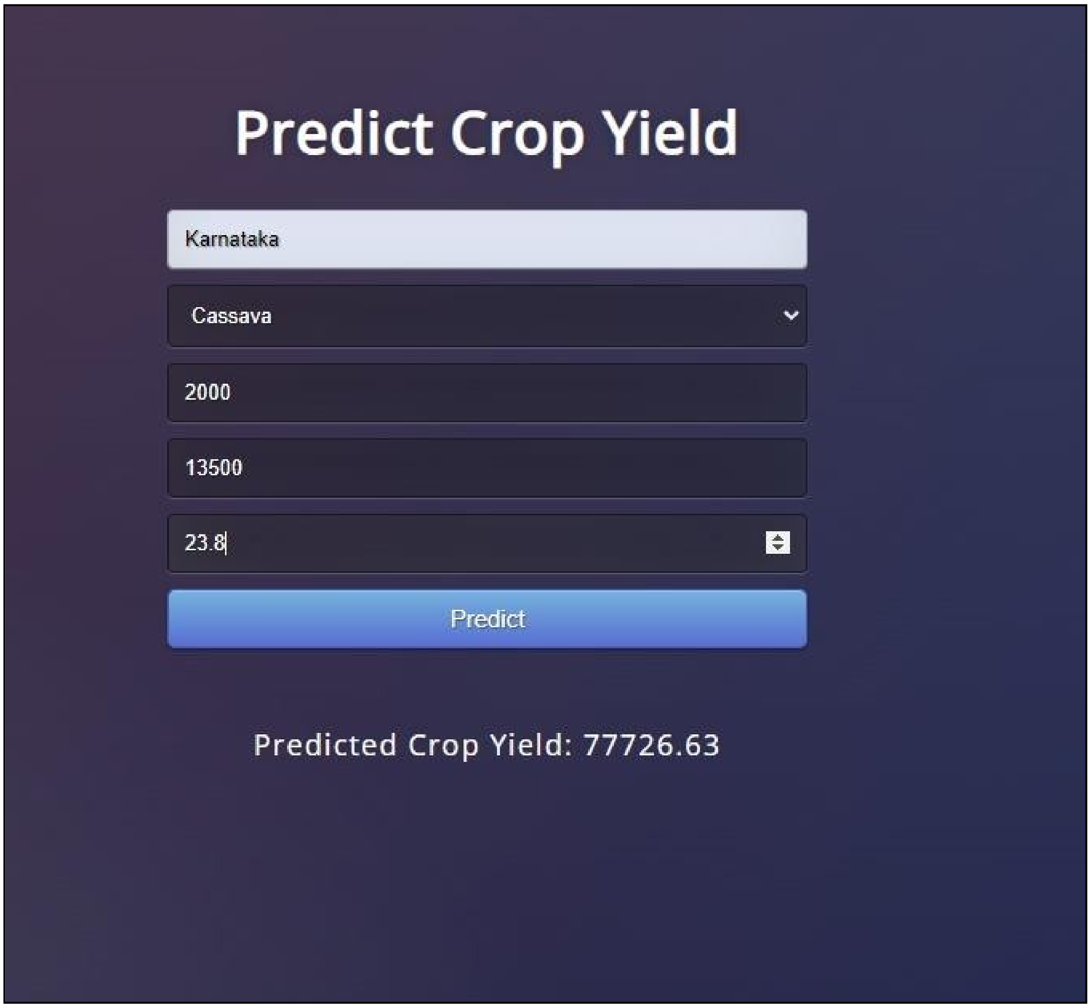


Predicting Values

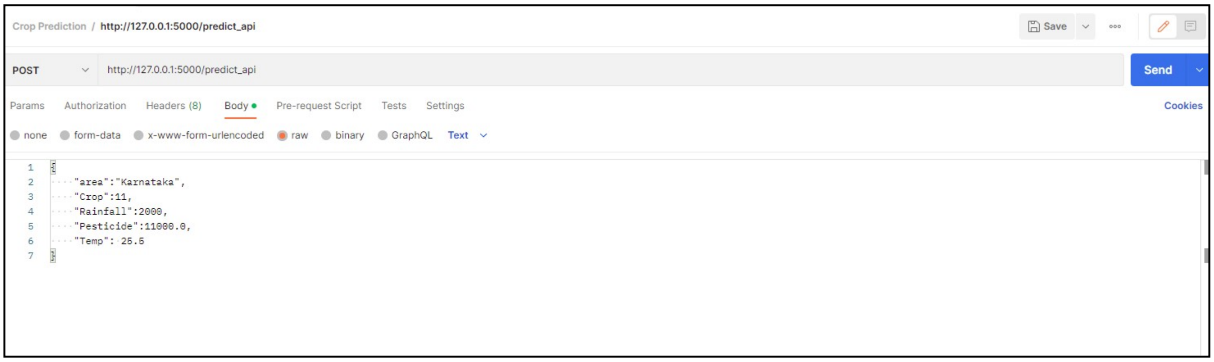
This is shown in the terminal/Jupyter Notebook

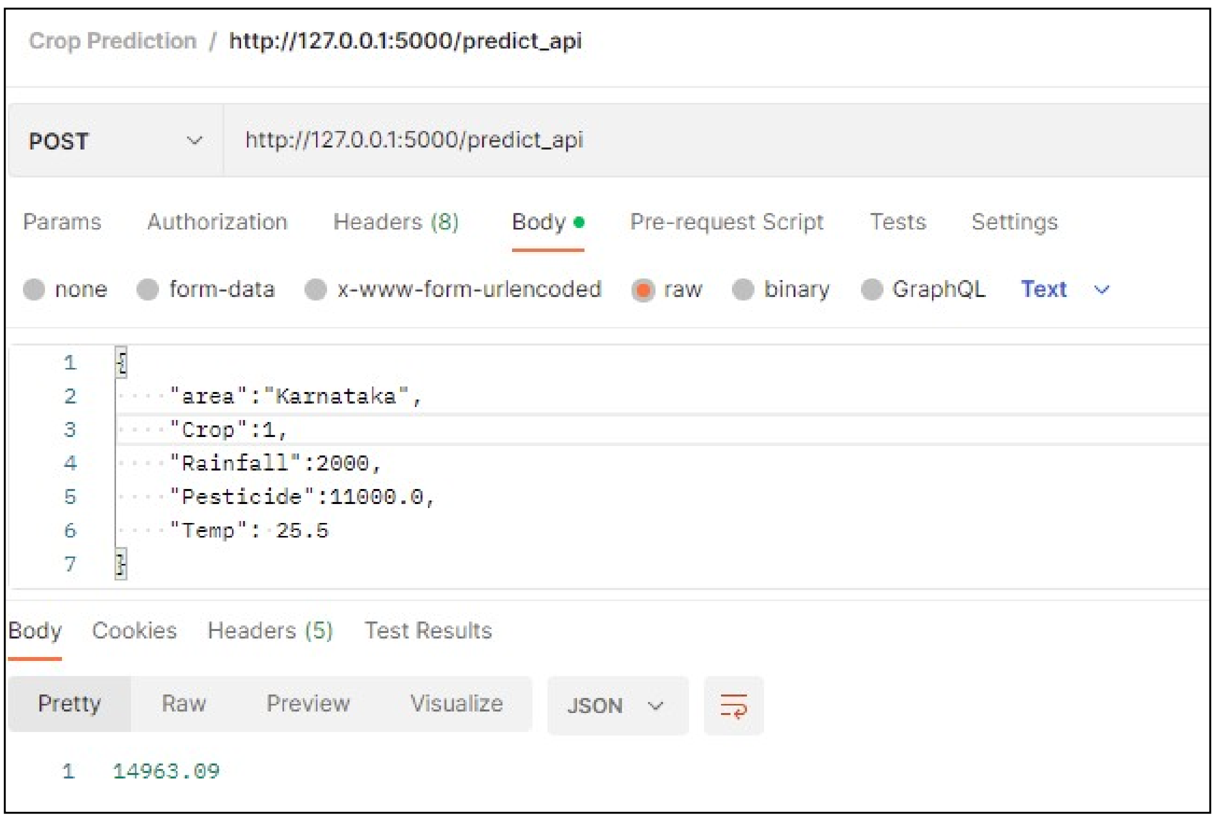
```
Enter the Indian State:Kerala
Please choose:
1) Cassava
2) Maize
3) Potatoes
4) Rice, paddy
5) Sorghum
6) Soybeans
7) Wheat
8) Sweet potatoes
9) Yams
Choose a Crop to predict Yield: 2
Enter the Year:2011
Enter average rainfall (in mm):1200
Enter amount of Pesticides (in tonnes):25000
Enter the average temperature (in Celcius): 23.22
Predicted hg/ha_yield is (Yield Per Hectare) is 18358.38
```

And this how prediction on our deployed app looks like



API Requests





6. Conclusion /Future Work:

Because our farmers are now not successfully utilising technology and analysis, there is a risk of incorrect crop selection for cultivation, which will diminish their income. To prevent these types of losses, we created a farmer-friendly algorithm that predicts which crop will produce the highest yield. As a result, farmers are more likely to make sound crop selection decisions, and the agricultural sector will benefit from innovative ideas.

We developed a crop yield prediction model that can estimate yield for eight distinct types of crops in India. According to our findings, the more datasets accessible, the more accurate the model. As the number of data points increases, our system will become more accurate. Our system is more accurate than the existing system. This model will assist farmers in growing crops that will produce more yield and hence be more profitable.

The constructed model has data points from 1990 to 2013, with 10 crops, 7 characteristics, and over 20,000+ records from the country acquired from a kaggle dataset. The dataset is specific to India and its states, and there are no missing numbers. Outlier detection and eradication were carried out. We use feature encoding and scaling for the string characteristics to do regression. The data was divided into 80% Train Data and 20% Test Data. A collection of Algorithms' accuracy score is examined. We determined that the Extra Tree Regression algorithm is most suited to developing the model for our dataset based on the average of accuracy scores for each technique, and we created our model accordingly.

In the future, this model can be applied globally by including data points from all countries. According to our investigation, as the number of data points rises, the accuracy of the model increases, therefore model data points can be increased to improve accuracy.

7. References:

[1] R. G. Durãaes, T. T. Salis, F. G. F. Coelho and A. d. P. Braga, "Explainability Analysis of a Machine Learning Model for Industrial Applications," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2022, pp. 1-6, doi: 10.1109/ICECET55527.2022.9872890.

[2] K. Ramu and K. Priyadarsini, "A Review on Crop Yield prediction Using Machine Learning Methods," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021, pp. 1239-1245, doi: 10.1109/ICOSEC51865.2021.9591764.

[3] S. S. Sajid, I. Huber, S. Archontoulis and G. Hu, "Integrating Crop Simulation and Machine Learning Models to Improve Crop Yield Prediction," 2022 17th Annual System of Systems Engineering Conference (SOSE), 2022, pp. 120-125, doi: 10.1109/SOSE55472.2022.9812678.

[4] Shidnal, Sushila & V Latte, Mrityunjaya & Kapoor, Ayush. (2019). Crop yield prediction: two-tiered machine learning model approach. International Journal of Information Technology. 13. 10.1007/s41870-019-00375-x.

[5] Wang, Wenguan & Shen, Jianbing & Ling, Haibin. (2018). A Deep Network Solution for Attention and Aesthetics Aware Photo Cropping. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2018.2840724.

[6] M. Kalimuthu, P. Vaishnavi and M. Kishore, "Crop Prediction using Machine Learning," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 926-932, doi: 10.1109/ICSSIT48917.2020.9214190.

[7] N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim and Z. Han, "Data Collection and Wireless Communication in Internet of Things (IoT) Using Economic Analysis and Pricing Models: A Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 4, pp. 2546-2590, Fourthquarter 2016, doi: 10.1109/COMST.2016.2582841.

[8] Shah, Ayush & Dubey, Akash & Hemnani, Vishesh & Gala, Divye & Kalbande, Dhananjay. (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. 10.1007/978-981-10-8339-6_6.

[9] N, Mahendra. (2020). Crop Prediction using Machine Learning Approaches. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS080029.

[10] Klompenburg, Thomas & Kassahun, Ayalew & Catal, Cagatay. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture. 177. 105709. 10.1016/j.compag.2020.105709.

[11] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2020, doi: 10.1109/ACCESS.2020.2992480.

[12] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin and N. Khan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction," in IEEE Access, vol. 9, pp. 63406-63439, 2021, doi: 10.1109/ACCESS.2021.3075159.

[13] D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using Machine Learning Algorithm," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1466-1470, doi: 10.1109/ICICCS51141.2021.9432236.

[14] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha and V. G. R. R. Devi, "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 736-741, doi: 10.1109/ICCES48766.2020.9137868.

[15] P. K. Gautam and S. Bhattacharjee, "Crop Yield Analysis using SIF and Climate Variables: A Case Study in Punjab, India," 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), 2022, pp. 148-154, doi: 10.1109/R10-HTC54060.2022.9929744.

[16] S. S. Sajid, I. Huber, S. Archontoulis and G. Hu, "Integrating Crop Simulation and Machine Learning Models to Improve Crop Yield Prediction," 2022 17th Annual System of Systems Engineering Conference (SOSE), 2022, pp. 120-125, doi: 10.1109/SOSE55472.2022.9812678.

[17] R. S. Renju, P. S. Deepthi and M. T. Chitra, "A Review of Crop Yield Prediction Strategies based on Machine Learning and Deep Learning," 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 2022, pp. 1-6, doi: 10.1109/IC3SIS54991.2022.9885325.

[18] N. Suresh et al., "Crop Yield Prediction Using Random Forest Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 279-282, doi: 10.1109/ICACCS51430.2021.9441871.

[19] P. Saini and B. Nagpal, "Deep-LSTM Model for Wheat Crop Yield Prediction in India," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), 2022, pp. 73-78, doi: 10.1109/CCiCT56684.2022.00025.

[20] A. Sharma, A. Tamrakar, S. Dewasi and N. S. Naik, "Early Prediction of Crop Yield in India using Machine Learning," 2022 IEEE Region 10 Symposium (TENSYMP), 2022, pp. 1-6, doi: 10.1109/TENSYMP54529.2022.9864490.