

---

# Crop Recommendation and Yield Estimation Using Machine Learning

---

A. Ashwitha<sup>1,\*</sup> and C. A. Latha<sup>2</sup>

<sup>1</sup>*Dept of ISE, MSRIT and Research Scholar-Dept of CSE, AMCEC, Bangalore, India and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*

<sup>2</sup>*Department of CSE, AMCEC, Bangalore, India and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*  
*E-mail: ashwitha.a.1990@gmail.com; drlathaca@gmail.com*

*\*Corresponding Author*

Received 18 September 2021; Accepted 05 December 2021;  
Publication 22 January 2022

## Abstract

In most developing countries like India, Agriculture is seen as one of the most widely followed habitations and contributes majorly to the country's economy. Agriculture provides the primary source of food, income, livelihood and employment to the majority of rural populations in India. Many crops are destroyed every year due to a lack of technical knowledge and unpredictable weather patterns such as temperature, rainfall, and other atmospheric parameters, which play a massive role in deciding the crop yield and profit. Therefore, choosing the right crop to grow and enhancing crop yield is an essential aspect of improving real-life farming scenarios. One of the motives is to collect and integrate the agricultural data from specific regions that may be used to analyse the optimal crop and estimate the crop yield. This script is novel by using simple crop, soil and weather parameters like crop, the area under cultivation, nitrogen, phosphorus and potassium content of the soil, season, average rainfall and temperature of a district in Karnataka, India. The user can predict the most suitable crop

*Journal of Mobile Multimedia, Vol. 18\_3, 861–884.*

doi: 10.13052/jmm1550-4646.18320

© 2022 River Publishers

and its estimated yield for a chosen year. This model uses primary classification, techniques like the random forest, k-NN, logistic regression, decision tree, XGBoost, SVM and gradient boosting classifier for determining the most suitable crop and regression algorithms like Linear Regression, k-NN, DBSCAN, Random Forest and ANN algorithm to estimate the yield of the most optimal crop identified earlier. The algorithm that has the least mean error is chosen for prediction and estimation and thus gives better results than the particular machine learning algorithm domain. There is a web interface for ease of use for end-users. Therefore, this project assists the farmers in choosing the suitable crop that can be grown in a particular region during a specific season or specific period and estimate its yield and predict if the recommended crop is profitable. Hence this project helps the farmers in preserving their time by assisting them in the decision-making process.

**Keywords:** Agriculture, machine learning, classification algorithms, regression algorithms, crop recommendation, yield estimation.

## 1 Introduction

Agriculture is the Indian economy's lifeline, as it employs around 70–75 per cent of the people directly or indirectly. Agricultural growth is closely proportional to India's economic growth. However, due to a variety of constraints, such as changing climatic conditions, bad soil conditions, soil erosion, land drifting, and floods, Indian farmers face various obstacles. If agriculture fails, it has a significant impact on the productivity of other industries, which harms the country's gross domestic product ratio. India's growth in agriculture has remained stagnant, and the government requires a technological revolution to fulfil the country's expanding population demands. Thus it is essential for us to adopt new technologies in agriculture and enhance crop production. This will help the farmers to avoid or minimise loss and achieve profit. Initially, the crop and yield predictions were performed based on farmer's experience in a particular region. And they don't have adequate knowledge on the changing weather patterns, soil components of phosphorus, potassium, nitrogen and the soil pH. Therefore this inadequate knowledge may lead to soil acidification, reduced yield, and damaging the top layer of the soil by the application of insufficient and improper amounts of nutrients to the land and without the crop rotation. The crop yield mainly depends on environmental changes, weather conditions, water resources, and rainfall.

Hence the farmers may not be able to achieve the expected or higher yield. Machine learning, data mining and deep learning are the essential analytical technologies that support accurate decision making in crop yield prediction, which includes some of the assisting conclusions on which crop to grow and the decisions regarding the crops in the growing season on the agricultural land.

### **Need for Prediction**

Food is one of humanity's most fundamental needs. The agricultural sector and its output determine it. In addition, the agrarian yield generation is strongly reliant on Indian food businesses. However, many farmers have committed suicide due to unfavourable climatic conditions and low yield output. Over 59,000 farmers are said to have attempted suicide in the last 30 years. Everyone nowadays is talking about agriculture and its production, but no one has come up with a methodology to assist them by increasing their crop productivity. A growing number of scholars are focusing their efforts on the agricultural industry. However, only a rudimentary attempt has been made in this regard. Motivation for the proposed system comes from the farming point of view that there is much work for the farmers to be done manually. This work focuses on algorithms that can forecast the most suited crop and estimate crop yield, assisting farmers in selecting and cultivating the most profitable crop, lowering the risk of loss and boosting the productivity and value of their agricultural area. Hence the proposed model assists the farmers in the decision-making process by recommending them the most optimal crop that, when grown, can produce greater yield so that with the help of technology, the agricultural sector in India will be progressed with innovative ideas. The past agricultural data is used in analysing crop production with the help of machine learning techniques, and it also provides a new way of understanding and analysing the various crops in the region of Karnataka and their crop productivity.

## **2 Literature Survey**

Various machine learning, data analytics, data mining, and deep learning techniques are previously used to analyse the agricultural data and draw valuable conclusions.

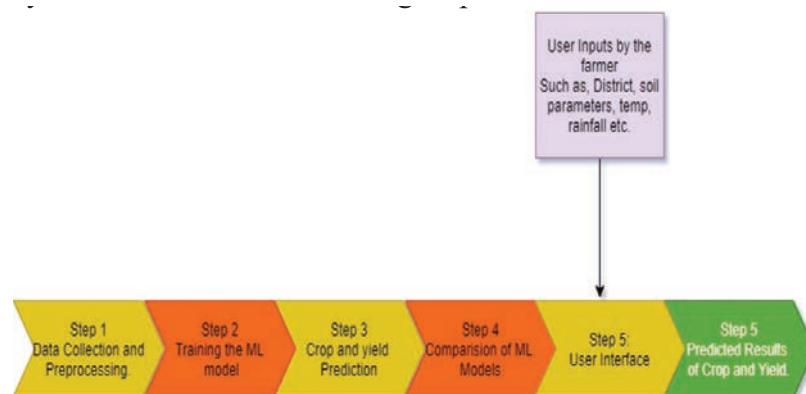
[1] have developed a web page using data mining to determine the crop yield on the basis of several climatic parameters as input. The C 4.5 algorithm

has been used to find the most predominant climatic input parameter on the yield of the crop in specific districts of Madhya Pradesh. The environmental data with respect to the previous years like temperature, rainfall were collected. It focuses on giving the idea of how the growth of the crop is affected by various input parameters of climate. [2] Developed a Java application to determine the crop yield rate by using two methods, namely, the K-nearest neighbours method and the Naive Bayes method. It helps in achieving the maximum rate of yield for the crops. It also assists in choosing an optimal crop suitable for a specific region and the environmental conditions. From the input dataset, the accuracy of the two methods is compared, and the best method is chosen for predicting. [3] proposed a system that uses data mining techniques to predict the harvest yield of the crop. This forecast helps the farmer to know his requirements and can plan accordingly. This system makes use of the Random forest algorithm for predicting the output of crop yield. [4] have analysed the agricultural data using data analytics and machine learning to assist the farmers by predicting the productivity of the crop in order to maximise crop production. They have made use of three algorithms, namely, 1. Naïve Bayes, 2. K-means Clustering 3. A priori Algorithm for final data accumulation and to predict the crop and its estimated yield. [5] have developed a model that uses the id3 algorithm for predicting crop yield and getting excellent quality and improved yield for the crop of Tomato. The proposed model is executed using PHP. The various parameters used in this study are Area, Temperature, humidity, and tomato crop production. [6] describes the rainfall and crop yield prediction using machine learning techniques. This paper describes the various machine learning techniques for predicting the rain and also predicting crop yield. It mentions the efficiency and the performance of various machine learning algorithms like SVM, decision tree, linear regression and KNN. It is concluded that the SVM algorithm has the highest efficiency or performs the best for rainfall prediction.

### **3 Proposed Approach**

This paper proposes a method for crop recommendation and yield estimation which is very important in the field of agriculture. Outcome can be visualised in the form of a user-friendly web interface.

The overview and the workflow of the project can be seen in Figure 1, which involves the steps involved and the sequential order that has to be



**Figure 1** Overview and analysis of proposed project.

performed in order to achieve the final results. The design of the proposed project of crop recommendation and yield estimation system involves the following steps:

- A. **Data Collection and integration:** It is the most crucial and essential step. The proposed model begins with the collection of the necessary agricultural data that consists of various parameters. It requires integrating data from multiple sources such as the soil parameters that include the nitrogen, phosphorus and potash values as are necessary for the crops, the weather parameters such as the season, rainfall, temperature, and Ph concerning a particular region and crops.
- B. **Data Pre-Processing:** The next step after data collection and integration is the data pre-processing, which also means cleaning the data and preparing it for modelling. The initial part involves handling the missing values, removing the noisy data, and handling the outliers. The next part is encoding the categorical values into numerical values using the label encoder to be easier in modelling the data. The next part is scaling or normalising the data by using the Minmax scalar or Standard scalar.
- C. **Splitting the data:** Once the data is pre-processed, the data is divided into two, i.e. the train data, which is used for training the model and the test dataset, which is used to test the model.
- D. **Applying Machine Learning Algorithms:** In this step, we apply various classification and regression machine learning algorithms and evaluate each model against the performance metrics.

**Crop Recommendation:** This module uses machine learning classification algorithms for modelling the data and recommending the suitable crop. The classification algorithms use are:

- Decision tree algorithm
- K-Nearest neighbours algorithm
- Random forest algorithm
- Logistic regression algorithm
- SVM algorithm
- XGBoost algorithm
- Gradient Boosting classifier
- Naive Bayes algorithm

**Yield Estimation:** This module uses machine learning regression algorithms to model the data and estimate the crop yield. The regression techniques used are:

- Linear Regression algorithm
- K-Nearest Neighbours algorithm
- Algorithm
- DBSCAN algorithm
- Random Forest algorithm
- Stochastic Gradient Descent Algorithm
- K means clustering
- Agglomerative Clustering

- E. **Data Exploration and Visualization:** This step involves exploring the data through visualisations to understand the properties and characteristics of the dataset and the results obtained in a better and easily understandable way.
- F. **Comparison of the Machine learning models:** The classification Models of crop recommendation and the regression models of yield estimation are compared and evaluated against various performance metrics. The crop recommendation models are compared with accuracy, precision and recall. The yield estimation models are assessed against mean absolute error. The algorithms that perform best in crop recommendation and yield estimation respectively are chosen.
- G. **User Interface:** The final output of the proposed project is a web application developed using a flask. The machine learning models of both crop recommendation and yield estimation are transformed into pickle models for integrating it with the flask application. The Web Application helps the users in accessing and using it efficiently. By

giving the required parameters for crop recommendation and yield estimation, the user will know the best crop suitable for a particular land and its estimated yield in an optimised manner.

## **4 Crop Prediction Algorithms**

The list of algorithms that compared used for crop prediction is as follows:

### **Decision Tree**

A decision tree functions as a decision support tool, with each branch representing a tree-like model decision. The decision tree results are insignificant for small datasets because of the number of options available at each branch, and the tree's height is easy to overfit [7, 10, 12, 17, 18, 21, 23–25].

### **Logistic Regression**

The logistic model is used in statistics to model the probability of a particular classification of an entity or an event occurring, such as fail or pass, win or lose, or positive or negative. This model can be transformed to replicate events, such as determining whether the image has a cow, buffalo, horse or pony. Each detected object in the image would be assigned a max probability of one, all of which adds up to one [14, 22].

### **Support Vector Machine**

Decision planes and boundaries are used in SVM. It divides the complete set of data into two categories. It works better when the data is multidimensional, and the number of dimensions exceeds the number of samples. When there is quite some separation between the classes, it works reasonably well. Feature scaling is required to achieve better outcomes. When the hyperplane has three or more properties, it is not easy to visualise [12, 23, 26].

### **XGBoost**

This is a decision tree algorithm that makes use of a gradient boosting framework. It can handle missing values. It has built-in regularisation, which prevents the model from overfitting. It is more likely to be overfitting than bagging. It requires a solid dataset and correctly chosen hyperparameters, which results in improved accuracy [22].

### **Gradient Boosting**

Gradient boosting is a technique that enhances the weak classifiers, like decision trees, for regression, classification, and other problems. A Gradient boosted decision tree can easily outperform a random forest in multiple scenarios. It carries out model construction in a stage-wise manner, similar to previous boosting approaches, then generalises them by allowing optimisation of any differentiable loss function [23].

### **Naïve Bayes**

The naive assumption taken for building this classifier is that the presence of one feature or attribute is independent of the existence of another set of traits or characteristics. In comparison to numerical variables, it works particularly well with categorical input variables. It can accurately estimate the class of a test dataset and operates quickly. The loss of categorical variables that were not observed in the training dataset might significantly impact the outcomes because it is foolish to assign zero probability to them. The underlying limitation is the assumption that the presence of one attribute is dependent on the presence of other qualities [12].

## **5 Yield Estimation Algorithms**

The list of algorithms that compared used for yield estimation is as follows:

### **Linear Regression**

The linear regression attempts to shape the relationship between two variables with the linear equation of the observed data. One variable is explanatory, while the other is considered to be dependent. For example, a person might want a linear regression model to match people's weight to their height [8, 14, 16, 19, 24, 26].

### **K-Nearest Neighbours**

In n-dimensional space, the clustering process is classified based on the k-nearest data points. The multi-class problem is simple to implement. It's straightforward to use. It is continually changing. The data quality determines the accuracy of the algorithm. And when the size grows, the calculations grow cubically. It is a computationally expensive algorithm [12, 15, 20, 22, 23].

### **Artificial Neural Network**

An ANN is built from a network of connected units or nodes known as artificial neurons, which are loosely modelled after the neurons in the human brain. Each link, like synapses in a human brain, can send a signal to other neurons. An artificial neuron that receives a signal analyses it and can signal neurons to which it is attached. The “signal” at a connection is a multiplier, and each neuron’s output is generated by the non-linear function of the sum of its inputs. The connection between neurons is referred to as edges. Neurons and edges usually have a weight that changes as learning progresses [11, 20, 22, 23].

### **DBSCAN**

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is used for spatial density-based clustering of data points in n-dimensional space. It is capable of handling clusters of diverse forms and sizes. It is noise-resistant. The DBSCAN algorithm is incompatible with data with variable spatial density. It also does not perform well with high-dimensional data [13].

### **Random Forest**

This is a classification algorithm that entails the construction of many decision trees. It reduces overfitting and enhances accuracy. It works effectively with both continuous and categorical data. The results are difficult to interpret because of the vast number of decision trees. When dealing with climatic variability, selecting the appropriate ranges of values to make decisions is critical [22, 23].

### **Stochastic Gradient Descent**

SGD is an optimisation approach used in Machine Learning applications to discover the best-fit attributes that predict the actual values. Because the network processes one of the data instances simultaneously, it simply lays through the memory. It is analytically faster because it only analyses one training sample at a time. Because of the large epoch values, the steps taken towards the minima are small and noisy. Furthermore, it takes longer to converge the loss function for obtaining the minima due to the noisy behaviour [22].

## Machine Learning Regression Algorithm

The user can choose out of two functionalities: crop prediction and yield estimation based on their requirement and enter the necessary inputs to obtain the desired results. Crop and yield recommendation is an approach that predicts the most optimal crop and the estimated yield by making use of various parameters such as temperature, rainfall, soil constituents and other environmental conditions. The main objective is to collect the agricultural dataset of a particular region, which can be used to analyse crop yield forecasting. The agrarian data is subject to several machine learning, data mining and deep learning techniques, and the comparison between these algorithms is made to recommend the crops more accurately. The agricultural data plays an important role, and using a large dataset with a more significant number of farming parameters, could help achieve better results. The farmer or the user of the application will have to enter the parameters such as selecting the district, choosing the season, Average temperature in Celsius value, NPK values, average rainfall and area of the field in hectare for the crop prediction and an added input of crop chosen for the yield prediction. After entering the information, the user will have to press the 'Predict crop' button on the crop prediction page and the 'Predict yield' button on the yield prediction page. The correct predictions will be made on the user's choice, and the result of the best crop which can be grown and predicted yield will be displayed in kilotons.

The pseudocode for the combination of results from the above set of chosen algorithms is as follows:

```

1: pred = None # store the predicted output
2: data = [] #store the input
3:
4: index()
5: # Renders home page
6:
7: predict(pred)
8: # Renders the crop prediction page.
9: if pred!=None:
10:     # Render page with result

```

```
11: else
12:     # Render page without results
13:     # Ask the users to fill the form
14:     input(data)
15:     submit()
16:     if valid (data) == True:
17:         # Predict crop
18:         pred = predict_crop(data)
19:         # Redirect with output.
20:         predict(pred)
21:     else
22:         # Show error message
23:     end if
24: end if
25:
26: yield_predict(pred)
27: # Renders the yield prediction page.
28: if pred!=None:
29:     # Render page with result
30: else
31:     # Render page without results
32:     # Ask the users to fill the form
33:     input(data)
34:     submit()
35:     if valid (data) == True:
36:         # Predict yield
37:         pred = predict_yield(data)
38:         # Redirect with output.
39:         yield_predict(pred)
40:     else
41:         # Show error message
42:     end if
43: end if
44:
45: about()
46: # Render about us Page
```

## Flowchart

The flow diagram of the Methodology is depicted in Figure 2 below.

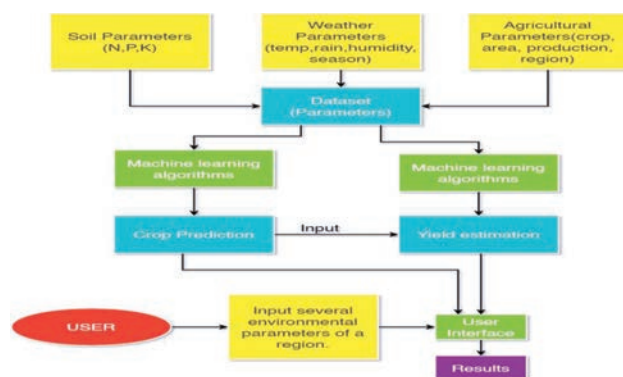


Figure 2 Flowchart.

## Results

### Data Collection and Integration

The data available in this work is obtained for the years from 1997 to 2014 in 30 districts of Karnataka in India.

A sample of integrated dataset can be shown in Figure 3.

State = Karnataka										
District	Year	Season	Area	Temp	pH	Rainfall	N	P	K	Crop
Udupi	2008	Rabi	39	25.15	5.3	123.21	45.3	8.1	6.2	maize
Chickballapura	2012	Kharif	134	26.32	6.5	82.34	56.8	9.4	4.5	Lobia
Bagalkot	2007	Kharif	1345	27.51	5.6	59.12	34.9	0	3.1	sunflower
Chickmagalur	2002	Kharif	3456	28.72	6.7	18.67	76.1	0	2.2	Tur
Bellary	2013	Kharif	7868	23.91	5.2	17.34	23.2	1.6	9.1	Tur
Dharwad	2001	whole	4565	34.22	6.3	34.15	59.3	7.8	5.3	Onion
Gulbarg	2007	Rabi	6786	27.43	5.4	23.34	20.4	9.0	8.5	sunflower
Bijapur	2012	Rabi	12	28.69	6.2	25.11	34.5	4.4	3.3	urad
Raichur	2011	Kharif	123	28.34	5.3	67.55	35.6	8.2	2.2	Mustard
Kolar	2002	Rabi	235	29.12	6.5	58.23	37.7	2.1	9.7	Mustard
Haveri	2002	Kharif	4345	30.23	5.8	23.34	53.2	0	1.	Sugarcane
Koppal	2007	Rabi	6778	26.44	6.9	37.19	41.3	0	4.3	moong
Bangalore	2007	Rabi	345	25.55	5.7	45.2	78.3	3.5	4.7	jowar
Hasan	2012	Kharif	3556	27.45	6.8	52.4	34.4	5.7	6.8	raagi
Mysore	1998	Rabi	768	28.31	5.8	98.8	23.9	8.5	7.6	rice
Belgaum	2000	Rabi	997	29.12	7.5	79.2	25.3	9.3	3.2	urad
Dharwad	2008	Kharif	675	24.23	6.9	83.2	37.1	7.2	2.1	linseed
Uttarkamataka	2003	Kharif	5678	23.34	5.2	62.3	59.2	6.2	1.2	moong
Koppal	2005	Rabi	3536	26.12	7.1	13.5	19.4	5.1	4.3	rice
kodagu	2011	Rabi	34346	28.67	5.4	92.8	28.7	0	8.3	gram
Yadgiri	2010	Kharif	3677	29.90	7.9	21.2	31.2	0	7.9	jowar

Figure 3 Sample data.

### Data Pre-Processing

The dataset collected had no null or missing, or duplicate values so, it did not require cleaning. After verifying this fact, the categorical are encoded into numerical values as shown in Figure 4.

State = Karnataka										
District	Year	Area	Production	Temp	pH	Rainfall	N	P	K	Crop
2	1997	11268	0.16	34.7	6.71	62.9	4.38	12.56	12.56	0
3	1997	10477	0.317	31.7	6.72	61.73	7.18	10.92	10.99	0
4	1997	6057	0.68	33.87	7.47	60.32	6.65	10.85	10.85	0
5	1997	48463	0.072	34.39	7.23	61.48	7.10	12.76	12.76	0
6	1997	20286	0.1196	33.04	7.19 2	64.18	4.02	12.69	12.69	0

Figure 4 Data after preprocessing.

### Splitting the Data

The collected dataset containing 6730 rows are split into test and train data. The test data has about 30% data or 2019 rows, whereas the training dataset includes the remaining 4711 rows.

The size vs data graph for the data split is depicted in Figure 5, which aids us in making appropriate analysis.

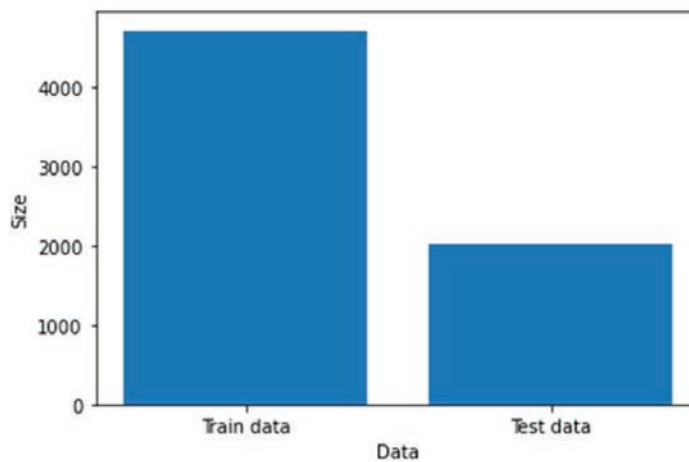
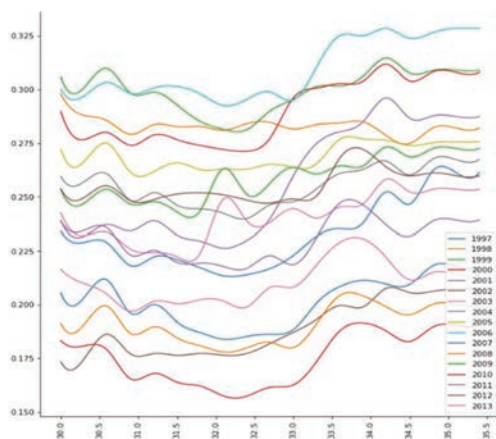


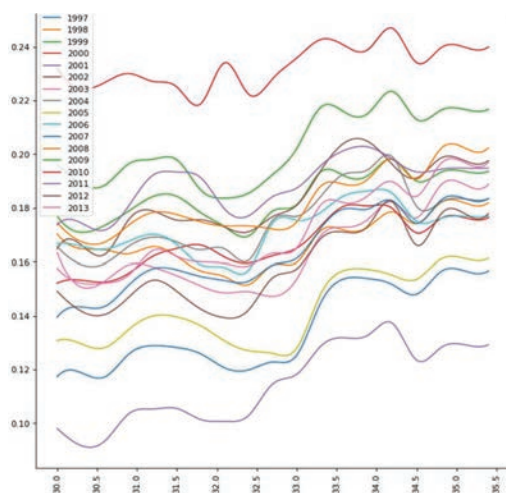
Figure 5 Plots of Data splitting for better visualisation.

### Data Analysis

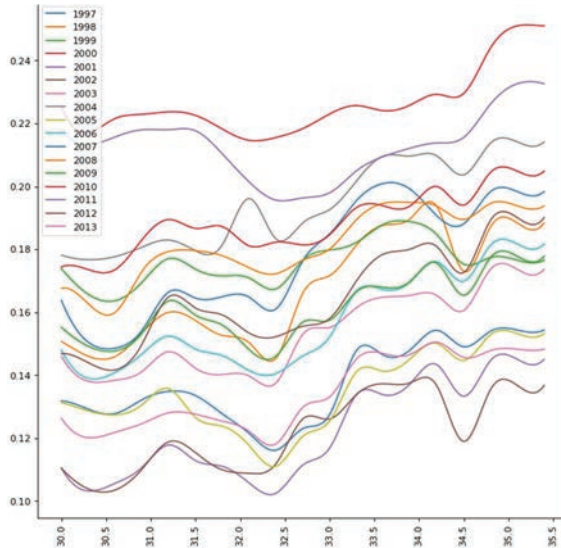
After analysing the graphs in Figures 6 to 8 which consists of Predicted yield vs Temperature for MOONG, we can identify that the optimal average annual temperature range for max yield for MOONG (GREEN GRAM) lies between 33 to 34.5 degrees Celsius. In contrast, the least favourable temp range is 31-to-32.5-degree Celsius.



**Figure 6** Predicted yield vs temperature over the years for MOONG (GREEN GRAM) in HASSAN.

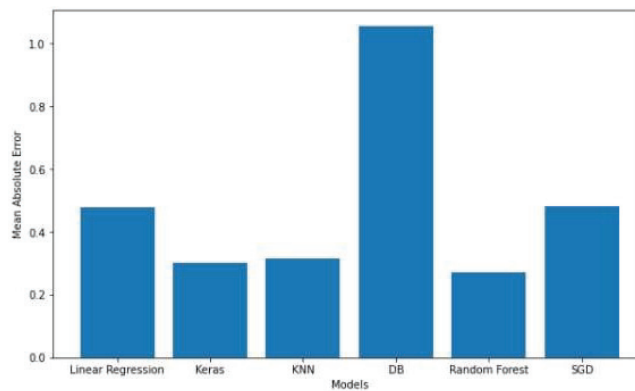


**Figure 7** Predicted yield vs temperature over the years for MOONG (GREEN GRAM) in BELGAUM.



**Figure 8** Predicted yield vs temperature over the years for MOONG (GREEN GRAM) in BIJAPUR.

The performance metrics are Accuracy for Crop prediction, whereas Root Means Square Error(RMSE) for Yield prediction. After comparing all the models as shown in the Figure 9, we concluded the Random Forest algorithm gave the highest accuracy among algorithms used for crop prediction and least RMSE among algorithms used for yield prediction. Random Forest achieved an accuracy of 99.94%. And RMSE of 0.2716.

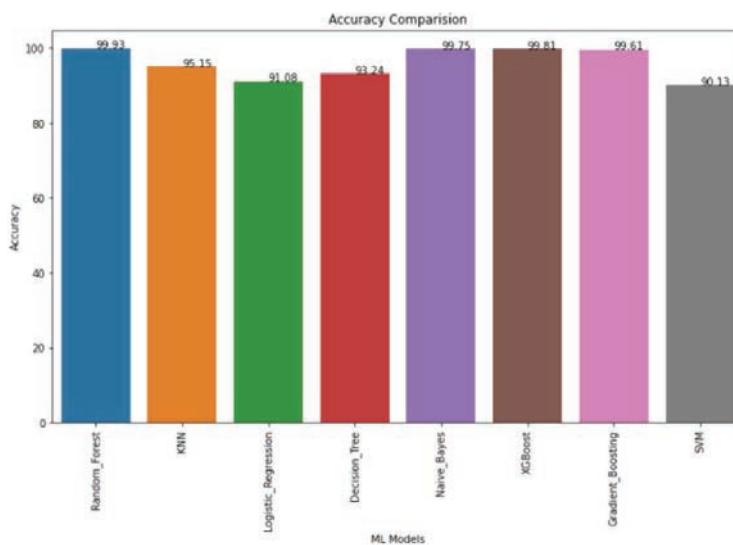


**Figure 9** The Mean absolute error comparison of the Machine learning models.

The Random Forest Regressor has the lowest mean absolute error of all the machine learning algorithms, as shown in Figure 9. As a result, it's best for predicting crop yield.

**Table 1** The mean absolute error comparison of the machine learning models for yield estimation

Machine Learning Models	Mean Absolute Error
Random Forest Regressor	0.270
Artificial Neural Net	0.3052
Linear Regression	0.478
KNN Regressor	0.303
DBSCAN	1.054
SGD Regressor	0.478



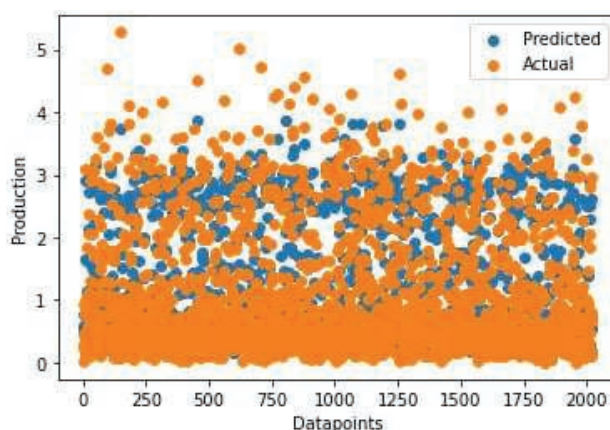
**Figure 10** Accuracy Comparison of the classification models of machine learning for Crop Recommendation.

Crop recommendation uses several classification algorithms, and each algorithm is evaluated and compared to find the best performing algorithm for the final prediction. From Figure 10, it is clear that Random Forest algorithm has the highest accuracy and it performs better than other algorithms.

The accuracy of all the models can be seen in the accuracy table in Table 2.

**Table 2** Accuracy of ML models for crop recommendation

Machine Learning Models	Accuracy
Random Forest	99.93
KNN	95.15
Logistic Regression	91.08
Decision Tree	93.24
Naive Bayes	99.75
XG Boost	99.81
Gradient Boosting	99.61
SVM	90.13



**Figure 11** The actual versus predicted yield of each crop.

In Figure 11, there is a close relationship between the actual and predicted data.

Crop recommendation and yield estimation features are included on a web page that is easy to access and use for farmers. Pickle is used to incorporating the algorithms chosen in the Flask application to develop the Web application based on selecting algorithms that perform better in crop suggestion and yield estimation.

The developed online application is divided into four web pages: home page, crop recommendation, yield estimation, and about us page.

**Crop Recommendation** – The user proposes the most appropriate crop and the projected return, as area selection, season entry, area values, temperature, precipitation, ph, nitrogen and potash, is also advised. The production output-input parameters are required. After the user enters suitable inputs, the predicted crop and estimated yield is displayed as shown in Figure 12.

Result!	
Perfect choice : maize-k	
Estimated yield : 1563.91	
District	MANDYA
Season	RABI
Area sown	200
Temperature	28
pH	6.3
Rainfall	85
Nitrogen (N)	50
Phosphorous (P)	20
Potash (K)	20
<input type="button" value="Predict Crop"/>	

Figure 12 Crop recommendation section of the web application.

**Yield Estimation** – Based on the input criteria District, season, rainfall, temperature, ph, crop, nitrogen, phosphorus, and potash, this component of the website determines the estimated yield of a specific crop. After the user enters suitable inputs, the estimated yield is displayed as shown in Figure 13.

Result!	
Estimated Yield : 1563.91	
District Name	MANDYA
Season	RABI
Crop Sown	MAIZE-K
Area sown	200
Temperature	28
pH	6.3
Rainfall	85
Nitrogen (N)	50
Phosphorous (P)	20
Potash (K)	20
<input type="button" value="Estimate Yield"/>	

Figure 13 Yield Estimation section of the web application.

## **Conclusion and Future Work**

Farmers' practises are among the most critical application areas in most developing countries, such as India. In agriculture, the use of analytical qualities of various technologies can affect the selection of multiple parameters, allowing farmers to build better and increased returns. Farmers are now unaware of the use of technology and the study of agricultural requirements, which may lead them to select fake plants for cultivation, affecting yields and thus income directly. A farmer-friendly system with a graphical user interface that recommends the most suitable and profitable crop for a particular site and is estimated in return is critical for preventing or minimising such loss. This proposed approach assists farmers in deciding on the best crop for farming, allowing the agriculture sector in India to progress through creative ideas and the use of technology. Past agricultural data is used for crop production research using machine learning algorithms, and it also presents a new approach to examining and analysing the varied cultures in the Karnataka region and their crop productivity. It is clear that the results of this method that different crops yield differently depending on precipitation, soil, temperature, season, and other factors. This technique allows farmers to recommend the best crop for production in the upcoming agricultural season and bridges the gap between agriculture and technology.

Crop analysis can be performed on a variety of other crops and in a variety of locations. Future IoT sensors and applications can be used to collect real-time agricultural farming data, and this data can also be integrated to prevent the use of harvest prices and crop fertiliser requirements with existing market data. Clustering techniques can also be combined to improve outcomes. Every land data relevant to this can be obtained by providing GPS sites for the specific country that can be accessed via the government's Rain Prediction System.

## **References**

- [1] Veenadhari, S., Bharat Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." *International Conference on Computer Communication and Informatics*. IEEE, 2014.
- [2] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques." *IEEE 5th International Conference for Convergence in Technology (I2CT)*. IEEE, 2019.

- [3] Kumar, Y. Jeevan Nagendra, et al. “Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector.” *5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2020.
- [4] Bhosale, Shreya V., et al. “Crop yield prediction using data analytics and hybrid approach.” *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, 2018.
- [5] CH. Vishnu Vardhanchowdary, Dr. K. Venkataramana, “Tomato Crop Yield Prediction using ID3.” *International Journal of Engineering Research and Technology (IJERT)* (Volume 4 Issue 10 pp. 663–62), 2018.
- [6] Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K T. “Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning.” *National Conference on Technology for Rural Development (NCTFRD-18)*, 2018.
- [7] Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. “Crop yield prediction using machine learning algorithms.” *Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125–130). IEEE, 2019.
- [8] Akshatha, Shailesh Shetty S, Anet P James, Athira M Saseendran, Chaitra M Poojary, “Crop Analysis and Profit Prediction using Data Mining Techniques” (Id:39), *International Journal of Engineering Research and Technology (IJERT) RTESIT – 2019 (VOLUME 7 – ISSUE 08)*.
- [9] Kamatchi, S. B., and Parvathi, R. (2019). Improvement of Crop Production Using Recommender System by Weather Forecasts. *Procedia Computer Science*, 165, 724–732.
- [10] Nishiba Kabeer, Dr Loganathan. D and Cowsalya. T. “Prediction of Crop Yield Using Data Mining.” *International Journal of Computer Science and Network (IJCSN)* (2019).
- [11] Talasila, Vamsidhar, Chitturi Prasad, Guttikonda Trinesh Sagar Reddy, and Allada Aparna. “Analysis and Prediction of Crop Production in Andhra Region Using Deep Convolutional Regression Network.” *International Journal of Intelligent Engineering and Systems (INASS)*, 2020.
- [12] Palanivel, K., and Surianarayanan, C. “An approach for the prediction of crop yield using machine learning and big data techniques.” *International Journal of Computer Engineering and Technology*, 10(3), 110–118.
- [13] Pandey A, Purohit S, Jadhav S, Shah K. “Optimum Crop Prediction using Data Mining and Machine Learning techniques.” *International*

*Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2014.

- [14] Rajeswari, S. R., et al. “Smart Farming Prediction Using Machine Learning.” *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 7 (2018).
- [15] Kusum Lata, and Bhushan Chaudhari. “Crop yield prediction using data mining techniques and machine learning models for decision support systems.” *International Journal of Emerging Technologies and Innovative Research (IJETIR)* (2019).
- [16] Devika, B., and B. Ananthi. “Analysis of crop yield prediction using data mining technique to predict the annual yield of major crops.” *International Research Journal of Engineering and Technology* 5.12 (2018): 1460–1465.
- [17] Mohan, P., and Patil, K. K. “Deep Learning-Based Weighted SOM to Forecast Weather and Crop Prediction for Agriculture Application.” *International Journal of Intelligent Engineering and Systems*, 11(4), 167–176 (2018).
- [18] Surya, P., and I. Laurence Aroquiaraj. “Crop Yield Prediction in Agriculture using Data Mining Predictive Analytic Techniques.” *IJRAR-International Journal of Research and Analytical Reviews (IJRAR)* 5.4 (2018): 783–787.
- [19] P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, “Crop Yield Prediction based on Indian Agriculture using Machine Learning,” *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020.
- [20] Sangeetha, C., and Sathyamoorthi, V. “Decision Support System for Agricultural Crop Prediction Using Machine Learning Techniques.” In *Proceedings of the International Conference on Intelligent Computing Systems (ICICS 2017–Dec 15th- 16th 2017) organised by Sona College of Technology*, Salem, Tamil Nadu, India.
- [21] Kumar, R., Singh, M. P., Kumar, P., and Singh, J. P. “Crop Selection Method to maximise crop yield rate using machine learning technique.” In *International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)* (pp. 138–145). IEEE, (2015, May).
- [22] Manoj G S, Prajwal G S, Ashoka U R, Prashant Krishna, Anitha P, 2020, “Prediction and Analysis of Crop Yield Using Machine Learning Techniques”, *International Journal of Engineering Research and Technology (IJERT)* NCAIT – 2020 (Volume 8 – Issue 15)

- [23] Rani, D. E., Satyanarayana, N., Vardhan, B. V., and Goud, O. S. C. "Crop Yield Analysis using Combinatorial Multivariate linear Regression." *International Journal of Advanced Research in Engineering and Technology (IJARET)* (2020).
- [24] Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese, and Er. Jinu Thomas. "Crop Prediction Using Machine Learning." *International Journal of Future Generation Communication and Networking* (2020).
- [25] Ms. Fathima, Ms Sowmya K, Ms Sunita Barker, Dr Sanjeev Kulkarni. "Analysis of crop yield prediction using data mining technique." *International Research Journal of Engineering and Technology (IRJET)* – 2020.
- [26] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate. "Crop Prediction Using Machine Learning." *International Research Journal of Engineering and Technology (IRJET)* (2020).

## Biographies



**A. Ashwitha** pursued Bachelor of Engineering from Visvesvaraya Technological University, Belgaum in 2011 and Master of Engineering from Bangalore University in the year 2013. She is currently pursuing a Ph.D. from Visvesvaraya Technological University under the guidance of Dr. C. A. Latha (Prof and Head of CSE, AMCEC, Bangalore) and working as an Assistant Professor in Department of Information Science, M S Ramaiah Institute of Technology. She had published 3 research papers in reputed international journals. Her main research work focuses on IoT Data Analytics, Data Mining and Computational Intelligence based education. She has 8 years of teaching experience and 4 years of Research Experience.



**C. A. Latha** pursued a Bachelor of Engineering from Mysore University in 1991 and a Master of Technology from NITK, Surathkal in the year 2003. She has pursued Ph.D. from Anna University, Chennai in the year 2012 and currently working as Professor and Head, Department of Computer Science, AMC College of Engineering, Bangalore since 2012. She is a member of Computer Society India since 1992 and a life member of the Indian Society for technical education since 1993. She has published more than 5 research papers in reputed international journals and conferences and authored a book, “programming in C and Introduction to Data structures”. She has reviewed several research papers for Elsevier and many international conferences and secured Best reviewer award in the year 2015 for her outstanding contribution in reviewing by Elsevier. Her main research work focuses on Cryptography Algorithms, Network Security.

