

# Crop Yield and Price Prediction System for Agriculture Application

Prameya R Hegde

Department of Computer Science and Engineering R V  
College of Engineering  
Bengaluru, India

**Abstract**—The rate of growth of agricultural output is gradually declining in recent years as the income derived from agricultural activities is not sufficient enough to meet the expenditure of the cultivators. The related factors responsible for the crisis include dependence on rainfall and climate, liberal import of agricultural products, reduction in agricultural subsidies, lack of easy credit to agriculture and dependency on money lenders, a decline in government investment in the agricultural sector, and conversion of agricultural land for alternative uses. As a predictive system is used in various applications such as healthcare, retail, education, government sectors, etc, its application in the agricultural area also has equal importance which is a statistical method that combines machine learning and data acquisition. In this project, the webpage is built using the Python Flask framework. Back end predictive model is designed using machine learning algorithms. developing a predictive model includes the collection of data, data cleaning, building a model, validation, and deployment. The aim is to provide a user-friendly interface for farmers and this model should predict crop yield and price value accurately for the provided real-time values.

## I. INTRODUCTION

Agriculture in India is a livelihood for a majority of the population and can never be underestimated as it employs more than 50% of the Indian workforce and contributed 17–18% to the country's GDP. Indian agriculture is characterized by Agro-ecological diversities in soil, rainfall, temperature, and cropping system. Monitoring crop growth and yield estimation are very important for the economic development of a nation. Many uncertain conditions such as climate changes, fluctuations in the market, flooding, etc, cause problems to the agricultural process. Technology can help farmers to produce more with the help of crop yield prediction.

Most of our Agricultural development programs in our country are mainly concentrated on providing resources and support after crop yields, there are no precautionary plans to make sure crop yields are obtained to full potential and plan crop cultivation. Nowadays, climate changes are predicted by the weather prediction system broadcasted to the people, but, in real-life scenarios, many farmers are unaware of this information. Crop yield estimation can be used to help farmers to reduce the loss of production under unsuitable conditions and increase production under suitable and favorable conditions. It also plays an essential role in decision-making at global, regional, and field levels

Ashok Kumar A R

Department of Computer Science and Engineering R  
V College of Engineering  
Bengaluru, India

In this research web-based application is built in which crop recommendation, yield prediction, and price prediction are introduced. This help the farmers to make better better management and economic decisions in growing crops.

## II. LITERATURE REVIEW

[1] According to analysis, the most used features are temperature, rainfall, and soil type, the most applied algorithm is Artificial Neural Networks in these models. According to analysis, Convolutional Neural Networks (CNN) are the most widely used deep learning algorithm in these studies, and the other widely used deep learning algorithms are Long-Short Term Memory (LSTM) and Deep Neural Networks (DNN). In [2] the prices of selected essential crops were analyzed for time-series prediction using meta-learning. Meta-Learning Based Adaptive Crop Price Prediction (MLACPP) is a combination of Self-Organized Map (SOM), LSTM (Long-Short Term Model), and this is used to train crop price datasets, and crop yield datasets. Experimental results show significant improvement in terms of prediction accuracy and cross-correlation entropy over the existing crop price prediction approaches.

In [3] Author used parameters like State, district, season, and area and the user can predict the yield of the crop in which year the user wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso, and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction. The performance metric used in this project is Root mean square error.

Paper [4] states that crop yield prediction incorporates forecasting the yield of the crop from past historical data which includes factors such as temperature, humidity, pH, rainfall, and crop name. The author used data mining techniques and random forest machine learning techniques for crop yield prediction. The proposed technique helps farmers to acquire apprehension in the requirement and price of different crops. It helps farmers in the decision-making of which crop to cultivate in the field.

In [5] paper the author proposes a forward feature selection in conjunction with hyperparameter tuning for training the random forest classifier. The accuracy of this method is 71.88%.

compared the accuracy of this method with two non-machine learning baselines. The first baseline used is the actual yield of the previous year as the prediction. The second baseline is that the target yield of each plot is manually predicted by a human expert.

In paper [6] Author states that Data mining and ML techniques can help to provide suggestions to the farmer regarding crop selection and the practices to get expected crop yield. Data mining uses the large historical data sets to create a new pattern to obtain the knowledge that helps in suggesting the farmers on selecting the crops depending on various available parameters and also helps in estimating the production of the crops. The author used the linear regression method to predict data also compared results with K Nearest Neighbor. The linear regression algorithm has proved more accurate prediction when compared with K-NN approach for selective crops.

In [7] Author states prediction of agriculture depends on parameters such as temperature, soil fertility, amount of water, water quality and seasons, crop price, etc. Machine learning plays an important role in crop yield prediction based on geography, climate details, and season. It helps farmers in growing the most appropriate crop for their farmland. The author used historical data and tested the prediction system for SVM (Support Vector Machine), random forest, and ID3(Iterative Dichotomiser 3) machine learning techniques. In terms of accuracy, SVM has outperformed other machine learning algorithms.

### III. METHODOLOGY

1) *Data Acquisition:* Three different types of data were gathered. The Dataset used for the experiment in this research is originally collected from the Kaggle repository and data.gov.in

#### a) Data for crop recommendation

This dataset helps to build a predictive model to recommend the most suitable crops to grow on a particular farm based on various parameters.

##### Context:

This dataset was built by augmenting datasets of rainfall, climate, and fertilizer data available for India.

##### Data fields:

N - the ratio of Nitrogen content in soil

P - the ratio of Phosphorous content in the soil

K - the ratio of Potassium content in soil

temperature - the temperature in degrees

Celsius humidity - relative humidity in %

ph - pH value of the soil

rainfall - rainfall in mm

#### b) Data for yield prediction

This dataset is a collection of crop yields from the years 1997 and 2018 for a better prediction and includes many climatic parameters which affect the crop yield

##### Data fields:

Corp Year: contains the data for the period 1997-2018

Agriculture season: contains all different agriculture seasons namely autumn, rabi, summer, Kharif, whole year

Corp name: contains a variety of crop names grown

primarily in India

Area of cultivation: In hectares

Temperature: temperature in degrees

Celsius Wind speed: In KMph

Pressure: In hPa

Humidity: In

percentage

Soil type: types found in India namely clay, loamy, sand, chalky, peaty, slit

Production: Produce per area in Tons

#### c) Data for crop price

This dataset contains all the geographical areas in India classified by state and district for the different types of crops that are produced in India from the period 2001-2015

Data fields: State. District, crop year, season, crop, and cost.

2) *Data pre-processing:* Three datasets that are collected are raw data that need to be processed before applying the ML algorithm. data collected are often incomplete, inconsistent, and lacking in certain behaviors or trends. They are also likely to contain many errors. So, once collected, they are pre-processed into a format the machine learning algorithm can use for the model. Used python pandas to visualization and analysis huge data. not required columns are removed. Empty columns are filled with mean values.

3) *Applying ML algorithm:* Some machine learning algorithm used are:

a) Decision Tree: It is a Supervised learning technique that can be used for both classification and Regression problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

b) Random forest: It is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

c) XGBoost: XGBoost is an implementation of Gradient Boosted decision trees. In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. The weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression,

classification, ranking, and user-defined prediction problems.

d) Lasso regression: It is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models.

e) Ridge regression: Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

Crop recommendation dataset consists of N, P, and K values mapped to suitable crops, which falls into a classification problem. Random forest classifier, XG boost classifier, and SVM are used to train the datasets and compare the result. For Yield, dataset output is a continuous value hence used random forest regression and ridge, lasso regression, are used to train the model. Similarly, for crop price prediction random forest regression, ridge and lasso regression is used to train. The algorithms for a particular dataset are selected based on the result obtained from the comparison of all the different types of ML algorithm.

#### IV. IMPLEMENTATION DETAILS

Crop recommendation, yield, and price data are gathered and pre-processed independently, after pre-processing, data sets are divided into train and test data. Data trained with ML algorithms and trained models are saved. Algorithms for a particular dataset are selected based on the result obtained from the comparison of all the different types of ML algorithms.

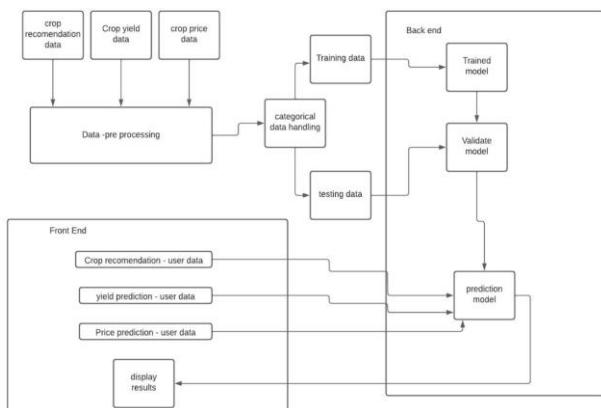


Fig. 1. System Architecture

The web interface is developed using flask, the front end is developed using HTML and CSS. Flask is a web framework that provides libraries to build lightweight web applications in python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself.

Users were able to enter the postal code and other inputs from the front end. Location and weather API is used to fetch weather data which is used as the input to the prediction model. Prediction models which deployed in back end makes prediction as per the inputs and returns values in the front end.

#### V. EXPERIMENTAL RESULTS

Crop recommendation is trained using SVM, random forest classifier XGboost classifier, and naive basis. It appears that the XGboost algorithm gives the highest accuracy of 95%. Crop yield and price prediction are trained using Regression algorithms. Random forest regression gives 92% and 91% of accuracy respectively. Detail comparison is shown in Table 1

The web application is built using python flask, HTML, and CSS code. It consists of sections for crop recommendation, yield prediction, and price prediction. Users can able to navigate through the web page and can get the prediction results.

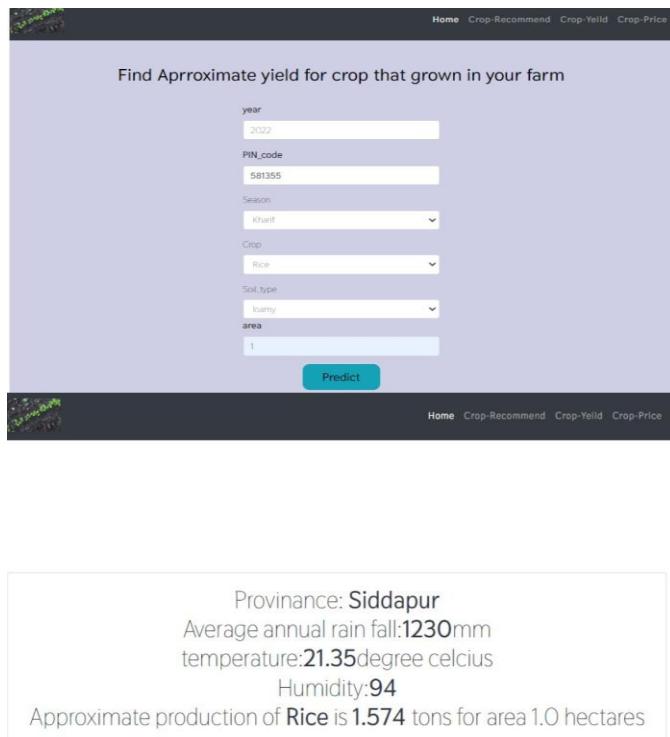


Fig. 2. The web interface of crop yield prediction

TABLE I  
 COMPARISON OF DIFFERENT ML ALGORITHMS ON  
 DATASETS

Data-set used	Algorithm	Accuracy	precision	recall
Crop recommendation data	Support Vector Machine (SVM)	92%	83%	88%
	Random forest classifier	90%	81%	87%
	Gaussian Naive Bayes	88%	90%	89%
	XGBoost	95%	89%	91%
Crop Yield data	Algorithm	Accuracy	Mean Absolute Error	Root Mean Squared Error
	Random forest regression	92.40%	81.2	367.574
	Ridge regression	70.60%	645.76	1229.66
	Lasso	75.57%	642.66	1229.18
Crop price data	Decision Tree Regressor	86.60%	105.101	470.38
	Algorithm	Accuracy	Mean Absolute Error	Root Mean Squared Error
	Random forest regression	91.40%	94.2	468.584
	Ridge regression	72.50%	685.78	1269.66
	Lasso	70.59%	678.66	1274.18
	DecisionTreeRegressor	80.60%	134.101	672.38

## VI. CONCLUSION AND FUTURE WORKS

This project must be able to develop a website where a Crop yield and price prediction model is deployed. The prediction system developed must take the inputs from the user and provide the best and most accurate predictive analysis for crop yield, and expected market price based on location, soil type, and other conditions. The website also provides information on the best crop that must be suitable for soil and weather conditions. The web page developed must be interactive enough to help out the farmers.

As a future scope, the web-based application can be made more user-friendly by targeting more populations by including all the different regional languages in the interface and providing a link to upload soil test reports instead of entering the test value manually.

## REFERENCES

[1] Thomas van Klompenburga, Ayalew Kassahuna, Cagatay Catalb “Crop yield prediction using machine learning: A systematic literature review” Computers and Electronics in Agriculture 177, 2020 Elsevier, DOI: 10.1016/j.compag.2020.105709.

[2] D. K. R. M., S. V. P. N., and I. A. Jayaraj, ”Meta-Learning Based Adaptive Crop Price Prediction for Agriculture Application,” in 2021 IEEE, 5th International Conference on Electronics, Communication, and Aerospace Technology (ICECA), 2021, pp. 396-402, DOI: 10.1109/ICECA52323.2021.9675891

[3] 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), 2020, pp. 1-4, DOI: 10.1109/INCET49848.2020.9154036

[4] 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

[5] Pausanias Charoen-Ung, Pradit Mittrapiyanuruk, ”Sugarcane Yield Grade Prediction Using Random Forest with Forward Feature Selection and Hyper-parameter Tuning” Recent Advances in Information and Communication Technology 2018 pp 33-42

[6] C. Chandana and G. Parthasarathy, ”A Comprehensive Survey of Classification Algorithms for Formulating Crop Yield Prediction Using Data Mining Techniques,” 2020 IEEE International Conference on Recent Trends in Information Technology (ICRTIT), 2020, pp. 1-6, DOI: 10.1109/ICRTIT51618.2020.92557403.

[7] G. S. Sajja, S. S. Jha, H. Mhamdi, M. Naved, S. Ray, and K. Phasinam, ”An Investigation on Crop Yield Prediction Using Machine Learning,” in 2021 IEEE, Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 916-921, DOI: 10.1109/ICIRCA51532.2021.9544815.

[8] 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/AC-CESS.2020.2992480.

[9] Li Tian, Chun Wang, Hailiang Li, Haitian Sun “Yield prediction model of rice and wheat crops based on ecological distance algorithm” in 2020 Elsevier, Environmental Technology & Innovation 20, DOI: 10.1016/j.eti.2020.101132

[10] 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.

[11] Y. Gadge and Sandhya, ”A study on various data mining techniques for crop yield prediction,” 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEEC-COT), 2017, pp. 420-423, doi: 10.1109/ICEEC-COT.2017.8284541.

[12] 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), 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154036.

International Conference on Technology, Engineering, Management for Crop yield and Price prediction System for Agriculture applicationSocietal impact using Market- ing, Entrepreneurship and Talent (TEMSMET), 2020, pp. 1-5, DOI: 10.1109/TEMSMET51618.2020.9557403.