National Institute of Technology, Patna

JAN 2022 - MAY 2022

The Crop Yield Prediction Using Machine Learning

By

Pranjal Mishra
Kshitij Khot
Priyanshu Bhardwaj

Under

Dr Anushree Tripathi
Department of Computer Science and Engineering
NIT PATNA

Abstract

In India, we all know that Agriculture is the backbone of the country. Agriculture is the first and foremost factor which is important for survival. Machine learning (ML) could be a crucial perspective for acquiring real-world and operative solutions for crop yield issues.
Considering the present system including manual counting, climate smart pest management and satellite imagery, the results obtained aren't really accurate.
The project focuses on predicting the crop yield in advance by analyzing factors like district state, season, crop type using various supervised machine learning techniques. This helps the farmers to know the crop yield in advance to plan and choose a crop that would give a better yield.

Keywords - agriculture; machine_learning; crop_yield_prediction; logistic_regression; naïve bayes; random forest; weather_api;

Introduction

Agriculture, since its invention and inception, has been the prime and pre-eminent activity of every culture and civilization throughout the history of mankind. It is not only an enormous aspect of the growing economy, but it's essential for us to survive. It's also an important sector for the Indian economy and also for the human future. It also contributes an outsized portion of employment. Because as the time passed the requirement for production has increased exponentially. So as to produce in mass quantities people are using technology in an exceedingly wrong way.
New sorts of hybrid varieties are produced day by day. However, these varieties don't provide the essential contents as naturally produced crops. These unnatural techniques spoil the soil. It all ends up in further environmental harm. Most of these unnatural techniques are wont to avoid losses. But when the producers of the crops know the accurate information on the crop yield minimizes the loss.
Machine learning, a fast-growing approach that's spreading out and helping every sector in making viable decisions to create the foremost of its applications. Most devices nowadays are facilitated by models being analyzed before deployment. The main concept is to increase the throughput of the agriculture sector with the Machine Learning models. Another factor that also affects the prediction is the amount of knowledge that's being given within the training period, as the number of parameters was higher comparatively. The core emphasis would be on precision agriculture, where So as to perform accurate prediction and stand on the various machine learning classifiers like Linear Regression, Decision tree, Random Forest etc. are applied to urge

a pattern. By applying the above machine learning classifiers, we came to the conclusion that the Random Forest algorithm provides the foremost accurate value. System predicts crop prediction from the gathering of past data. Using past information on weather, temperature and a number of other factors the information is given. The Application which we developed, runs the algorithm and shows the list of crops suitable for entering data with predicted yield value.

India being a developing nation with a rapidly growing population, food consumption is high and will increment in the near future. Thus to establish food security, vertical advancement in agribusiness is of great importance.. Likewise, it turns out to be fundamental to replicate and forecast the crop yield under encompassing conditions, prior to the application phase for efficient crop management. Since the association between crop yield and climatic and non-climatic parameters are nonlinear and incorporate some difficulties, machine learning may result in a rewarding option for crop yield prediction.

Machine learning administers specific strategies to characterize principles and patterns in extensive datasets relevant to crop yield with eminent forecasting capacity. Also, it would be able to improvise the forecasting model. At present agricultural advances are largely redirected to machine learning algorithms since it has augmented crop yield with limiting information cost. Machine learning algorithms enhance the agriculturists to improve the crop selection and crop yield forecast.

Now we are going to focus on the practical implementation of the solution.

Literature survey

In many of the research papers it has been found that everyone uses climatic factors like rainfall, sunlight and agricultural factors like soil type, nutrients possessed by the soil (Nitrogen, Potassium, etc.) but the problem is we need to gather the data and then a third party does this prediction and then it is explained to the farmer and this takes a lot of effort for the farmer and he doesn't understand the science behind these factors. To make it simple and which can be directly used by the farmer, this paper uses simple factors like which state and district the farmer is from, which crop and in what season (as in Kharif, Rabi, etc.)[1]. Some people conducted experiments on Indian government dataset and it's been established that Random Forest machine learning algorithm gives the best yield prediction accuracy. Results reveal that Random Forest is the best classifier when the relevant parameters are combined[5].

have implemented crop yield prediction by using only the random forest classifier. Various features like rainfall, temperature and season were taken into account to predict the crop yield. Other machine learning algorithms were not applied to the datasets. With the absence of other algorithms, comparison and quantification were missing thus unable to provide the apt algorithm[6].

In all the reference papers referred, we found that most of the prediction have been performed on the basis of climatic factors

Methodology

Through The crop yield prediction system better planning and decisions can be made by the farmer before sowing the seeds. This predictive Modal has been built as follows:

1. Data Preprocessing and Feature selection

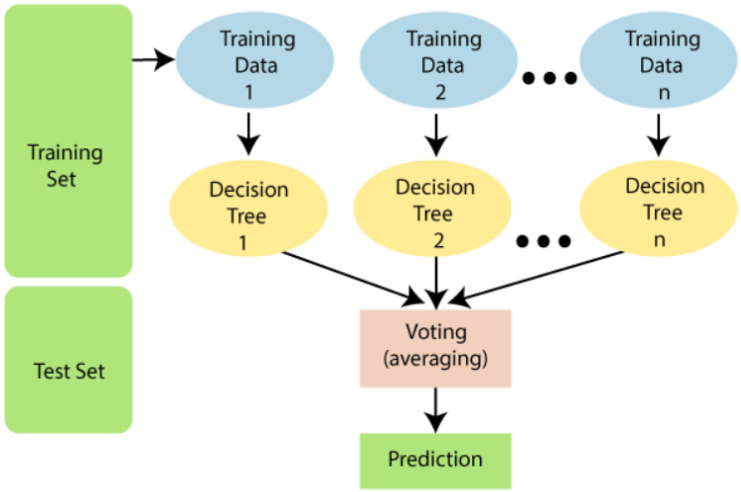
The dataset collected from the Indian Government Repository is first cleaned as It may contain some incomplete, redundant, inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized. After that the irrelevant data which is not a feature for our model should be dropped and finally we are left with the feature selected dataset. This includes factors affecting yield and production.

2. Algorithm Selection

*Linear Regressio*n - Linear regression is used to predict the relationship between two variables by applying a linear equation to observed data. There are two types of variable, one variable is called an independent variable, and the other is a dependent variable. Linear regression is commonly used for predictive analysis. It takes into account the significance of independent variables and predicts the dependent variable.

Decision Tree - Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

Random Forest - Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.



There are a number of key advantages and challenges that the random forest algorithm presents when used for classification or regression problems. Some of them include:

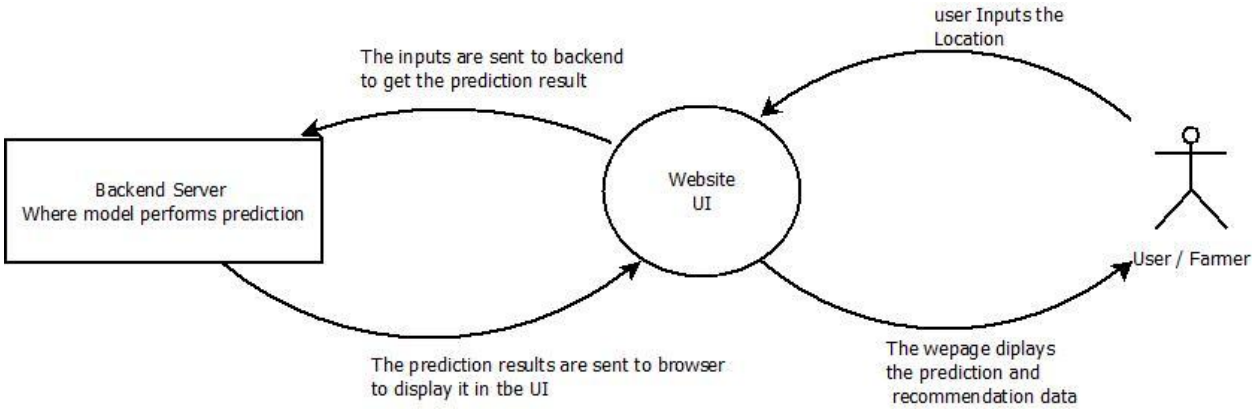
- Reduced Risk of overfitting
- Provides flexibility
- Easy to determine feature importance

Finally, On implementing the above three algorithms we come to the conclusion that our model is most accurate in case of Random Forest Algorithm. Thus it can be used to train the model.

Algorithm	Accuracy
Linear Regression	20.3545 %
Decision Tree	92.1678 %
Random Forest	97.3874 %

3. Software Architecture

Our Crop yield prediction consists of both frontend and Backend. When the user Inputs the location the data is set to the backend and the prediction model predicts. The prediction is sent to the client / browser where the react framework is used to display the prediction results.



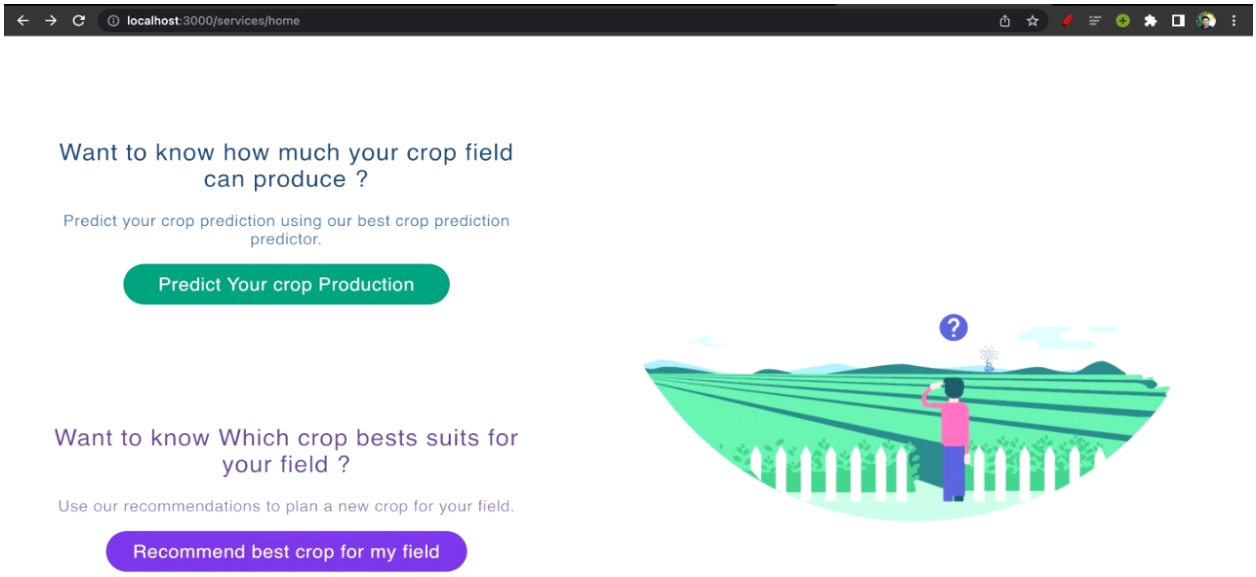
Proposed System

The crop yield prediction system is going to be a very handy tool for agriculture-related users. The system will let the user input the location (i.e. city, state, etc) and on the basis of it the modal will predict the yield for that location. Also the system is made to recommend the most productive crops for the location

Results

As we have decided to use better and more efficient methods, our accuracy would be far better than the earlier works.

A react web-platform is developed for this proposed work which has a user-friendly interface that can be accessed by the user. . Once the location-related data is given as input region & season is fetched and compared with the crop dataset, a list of crops suitable for given input condition will be displayed.



Conclusions

This work presents an effective crop prediction system by using past data. Using Machine Learning model and functions datasets are analyzed and the trained model is used to fetch the crop based on region and season. This proposed system helps the farmers to select suitable crops based on season and region of sowing. It will in- turn help the farmers by reducing the loss faced by them and improve net crop yield.
This also contributes to our Indian economy by increasing the yield rate of crop production.

References

1. *Crop Yield Prediction based on Indian Agriculture using Machine Learning - 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020*
2. *A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters - Dhivya Elavarasan, · P. M. Durai Raj Vincent, Journal of Ambient Intelligence and Humanized Computing (2021)*
3. *Fuzzy deep learning-based crop yield prediction model for sustainable agronomical frameworks - Dhivya Elavarasan, · P. M. Durai Raj Vincent, Neural Computing and Applications (2021)*

4. *Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey - Nishu Bali, Anshu Singla, Archives of Computational Methods in Engineering (2022)*
5. *Crop Yield Prediction Using Machine Learning Algorithms - Aruvansh Nigam, Saksham Garg , Archit Agrawal, Parul Agrawal, 2019 Fifth International Conference on Image Information Processing (ICIIP)*
6. *A Study on Various Data Mining Techniques for Crop Yield Prediction - Yogesh Gandge, Sandhya, 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*
7. *Crop Prediction using Machine Learning - M.Kalimuthu, P.Vaishnavi, M.Kishore, Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICCSIT 2020)*
8. *An efficient algorithm for predicting crop using historical data and pattern matching technique Anjana, Aishwarya Kedlaya K , Aysha Sana, B Apoorva Bhat, Sharath Kumar, Nagaraj Bhat*
9. *Crop Yield Prediction using Machine Learning Algorithms - Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Prof. Vinu Williams Department of Computer Science and Engineering*
10. *https://www.data.gov.in*
11. *Machine Learning Random Forest Algorithm - Javatpoint*