

# **Crop Yield Prediction using Machine Learning**

## **Motivation and Problem Statement**

Machine learning is an important decision support tool for crop yield prediction, including supporting decisions on what crops to grow and what to do during the growing season of the crops. Several machine learning algorithms have been applied to support crop yield prediction research.

Crop yield prediction is one of the challenging problems in precision agriculture, and many models have been proposed and validated so far. This problem requires the use of several datasets since crop yield depends on many different factors such as climate, weather, soil, use of fertilizer, pH and seed variety.

This indicates that crop yield prediction is not a trivial task; instead, it consists of several complicated steps. Nowadays, crop yield prediction models can estimate the actual yield reasonably, but a better performance in yield prediction is still desirable. So, the models need to be trained using datasets, where the outcomes are represented based on past experience. The predictive model is built using several features, and as such, parameters of the models are determined using historical data during the training phase. For the testing phase, part of the historical data that has not been used for training is used for the performance evaluation purpose.

You need to create a ML Model that takes all the above factors in consideration and give predictions/prescriptions that would help farmers

to increase their crop productivity, and hence gain more profit by being more productive.

## **Dataset Overview**

Given below are the links to the datasets to be used for the project.

1.State Wise Major Crop Production:

<https://data.world/thatzprem/agriculture-india>

2.Crop Dependence on Various Factors:

<https://drive.google.com/file/d/1QP2ZVWSjtqRk7R4oCzcA9dtTBUm0wlvw/view?usp=sharing>

3. Ideal value of pH for various crops

<https://drive.google.com/file/d/1T3BuaUu7eVZ4lidR3RgMDSbtUgPjIMGv/view?usp=sharing>

4. State Wise Cost of Production

[https://drive.google.com/file/d/1q9NBmOI5PKIYYshuYJlbFg717BPTS1\\_a/view?usp=sharing](https://drive.google.com/file/d/1q9NBmOI5PKIYYshuYJlbFg717BPTS1_a/view?usp=sharing)

5. Actual Crop Yield based on Various Factors

[https://drive.google.com/file/d/1j2PFa0czTiE-mxcA5D3y5SX6XMYnwt\\_F/view?usp=sharing](https://drive.google.com/file/d/1j2PFa0czTiE-mxcA5D3y5SX6XMYnwt_F/view?usp=sharing)

## **TASKS**

1. Perform an EDA on the above datasets to explore and find the best crop to grow in each State. Also, list out the causes that make them

most suitable for growing in those regions. And what factors affect the respective crops the most.

2. The above datasets provide a huge amount of information on crop production in India ranging from several years. Based on the Information the ultimate goal would be to predict crop production using powerful machine learning techniques. But before you can build any model, you need to collect the relevant data in one place. There are various datasets provided in the above link, but you might need to do some preprocessing on them to make them suitable for any analysis, or building your machine learning model.
3. Create an interface(using statements), where the user inputs their Geographic Location(District), and based on that your model should generate output prescribing them the best crop to grow in that region.
4. For more accurate prescription, collect more metrics(features) from the user on which the crop yield is dependent, and then predict the crop to grow, along with the expected profit from that crop.
5. If any of the metrics is not available then the model should be able to handle such scenarios and predict the outcomes nonetheless.
6. The final output should be in the form of a dataframe where columns should include the current crop the user is growing, the suggested crop, current profit from the present crop, expected profit from the predicted crop.
7. If the information about soil health is available, then suggest the prescriptive actions for the soil improvement if any, and the cost associated with the change.

8. Experiment with various models for this project. And select the best one that fits the best in this situation. Give detailed analysis for your selection.
9. You may use visual tools to present your findings keeping in mind your audience.