CROP PREDICTION USING MACHINE LEARNING-A STEP TOWARDS SELF RELIANCE

—

Smriti Sharma
MTECH. - AI and DS
01902212022

Overview

India is considered as a global agricultural powerhouse by the World. India holds the second largest agriculture land in the world with approximately 179.9 million hectares under cultivation. It is the second largest producer of rice, wheat, cotton, sugarcane, fruits, vegetables and tea.

The Agricultural Sector plays a very important role in the Indian Economy , contributing about 18.8% (FY 2021-22) to the country's GDP . A major strata of India's Population rely on Agriculture for their livelihood. It has been surveyed that approximately 54.6% of India's population is engaged in agriculture.

As the population of India is increasing , so is the demand for food .

Since the Indian farmers are not educated well and afraid of cultivating different crops, crop diversification is the biggest challenge. We still are the largest importer of food products. The farmers have been cultivating the same crops for generations and have a belief in the myth that using more fertilizers and pesticides will yield more production. Instead of soil analysis and cultivating the environmentally suitable diversified crops which will not only help them to earn more but also will meet the diversified food demands of the Indian Population. Many Indian farmers are still using irrationally high quantities of Fertilizers and Chemicals in fields which adversely affects soil, ground water and quality of crops.

Problem Statement

With the increased use of fertilizers, pesticides and other chemicals by farmers for maximizing the yield of crops within a shorter span of time in order to earn more profit results in the production of very high chemically contaminated crops of grains ,fruits and vegetables . These are considered to be one of the key contributors behind various life threatening diseases in human beings and animals. The use of these chemicals/pesticides not only affects humans and animals but also has a significant effect on the environment particularly Groundwater.The use of chemicals/pesticides results in contamination of Groundwater, which will affect the generations to come.

Moreover the high use of fertilizer and pesticides adds to the pocket of farmers.Taking this phase forward and adding to sustainable farming, Our project "Crop Prediction Using Machine Learning- A Step Towards Self Reliance " aims at :-

1. Limiting the use of fertilizers and pesticides while farming ,
2. Increasing the income of the farmers, by providing them with appropriate crops that are suitable to be grown on their land, depending upon the soil analysis (Crop Prediction) and also providing them about the overall profitability in the entire process (Price Prediction).

This would benefit us in the following manner:-

1. The quality of the crop is maintained automatically , as that crop is grown in the most favorable soil condition based on soil analysis which leads to very limited and restricted use of fertilizers and pesticides.
2. It will yield more profits to the farmer as the cost of buying fertilizers and pesticides is restricted and the labor cost will be reduced significantly.
3. It would also add to the efforts of the government towards the diversification of crops and increase the income of the farmers.
4. Restricted / Limited use of fertilizers and pesticides will not adversely affect the environment, particularly quality of groundwater.
5. Dependability on import will reduce significantly and would help India in achieving the Goal of Self Reliance in the agricultural Sector.
6. Diversified Crop Production, High Yield of crop per acre, Export Quality crops will add up in India's GDP.

Datasets

1. The dataset used for the research is the Crop Recommendation Dataset which is a class balanced dataset. The dataset consists of different varieties of crops such as wheat ,rice, maize etc. which require different soil and weather conditions for cultivation. The size of the dataset was 2200 rows and 8 columns , the dataset was last updated in 2020 .The dataset consists various features such as (N) ratio of Nitrogen content in the soil, (P) ratio of phosphorus content in the soil, (K) ratio of potassium content in the soil, temperature in degree celsius ,relative humidity in the atmosphere in(percent) , ph value of the soil, amount of rainfall in mm and the target column which is the variety of crop.

   In order to widen up the area of our research we added a few more varieties of crops to our dataset thus increasing the size of the dataset to 3100 rows and 8 columns .The data for (N, P, K , Temperature, Humidity, Ph, Rainfall) for the newly added crops are collected by referring to various Government websites such as Farmer's Portal.gov.in and Crop Cultivation Guide.

Taking into consideration soil analysis and weather conditions. This dataset helps us to predict the most suitable crop , which is favorable to be cultivated under the given conditions.

LINK- CROP RECOMMENDATION DATASET| KAGGLE

2. The second dataset taken into consideration for the purpose of price prediction is Agri Commodity Min, Max, Modal Price Dataset. The dataset was last updated in 2021. The size of the dataset is (62.31MB). It consists of 348 Commodities prices market wise. This dataset consists of the Name of the Commodity ,State , District, Market, Minimum Price (Quintal => Kg Converted) ,Maximum Price (Quintal => Kg Converted) , Modal Price (Quintal => Kg Converted), Type of Crop,Date on which the data is obtained

   LINK-🌾 AGRI COMMODITY MIN , MAX ,MODAL PRICE DATA | KAGGLE

Data Cleaning

Data preparation:

- **In Excel**

  Values of N-P-K of a few crops are added.

- **In Python**

  Data is uploaded via the Pandas read_csv() function and returns a new Dataframe with the data and labels from the file data. csv.
  NULL values are identified for the value column and rows containing NaN in the value columns are dropped.