Crop Yield Prediction Using Decision Tree

Aleena Joseph
PG Scholar
Master of Computer Applications
Amal Jyothi College of Engineering
Kottayam, Kerala
aleenarejijoseph@gmail.com

Ajith G S
Assistant Professor
Master of Computer Applications
Amal Jyothi College of Engineering
Kottayam, Kerala
gsajith@amaljyothi.ac.in

***Abstract*— Machine learning-based agricultural yield prediction is a crucial tool for farmers, assisting them in making knowledgeable crop selection. One of the primary objectives of predicting crop yields is to support farmers in choosing the most appropriate crops by considering climatic conditions and enhancing agricultural methods. These models are made to assist farmers in maximizing crop output by determining the ideal cultivation settings. The models consider factors such as weather conditions, pesticides, and Area to make the predictions. Decision Tree models are an important tool in agriculture because research has shown that they can achieve high accuracy rates. These models can help farmers understand which crops are best for their particular conditions and improve their farming methods.**

Keyword—Agriculture, Crop Yield Prediction, Decision Tree

I. INTRODUCTION

Agriculture is a vital sector for economies worldwide, playing a critical role in food production and resource provision. Crop yield is the amount of agricultural products produced from a specific plot of farmed land, usually expressed in terms of weight or quantity. Crop variety, soil quality, farming methods, and climate conditions are just a few of the many variables that can affect crop yield. To forecast crop yields, technologies like machine learning and data analytics have been used.

Accurate forecasts of crop yield empower farmers to make knowledgeable choices regarding crop choice, planting schedules, and resource distribution, thereby optimizing their farming methodologies, among other aspects. Enhanced sales of agricultural produce resulting from abundant crop yields play a pivotal role in fostering economic steadiness. The utilization of diverse machine learning algorithms can be instrumental in forecasting crop yields. A number of algorithms, including the K Nearest Neighbor (KNN), Decision Tree, Random Forest, and others, are used to forecast crop yields. The Decision Tree Algorithm is employed in this work. This approach underscores the flexibility and versatility of machine learning in agriculture, offering farmers valuable insights into future crop yields and aiding in better decision-making for sustainable farming practices. Here the prediction is done based on the factors such as the Climatic Conditions, Area, Crop, and the Pesticides used. By using this the crop yield predictions can be made more precisely.

The proposed models are briefly explained in the parts that follow, together with a general architecture. So this prediction will helps to predict the crop yield of a specified area.

II. RELATED WORKS

A. In the article [1] paper employs advanced regression techniques, including Kernel Ridge, Lasso, and ENet algorithms, to forecast crop yields. Moreover, it leverages the concept of stacked regression, enhancing the effectiveness of these algorithms and ultimately delivering more precise yield predictions.

B. In [2] Using artificial intelligence and machine learning techniques, the paper conducts a detailed analysis of various aspects associated with crop yield prediction in India. In order to analyze a wide range of variables, including soil conditions, climate, historical yield data, and agricultural practices, this research uses sophisticated computational methods. The study aims to improve crop yield forecast accuracy by applying AI and machine learning algorithms, empowering Indian farmers to make informed decisions about agricultural strategies and crop selection.

C. The researchers in this paper[3] utilized a variety of machine learning methods to anticipate crop yields for five distinct crops grown in the Indian state of Rajasthan. Amid the assortment of techniques employed, the Random Forest regression method exhibited exceptional performance, surpassing other methods such as Support Vector Machine (SVM), Gradient Descent, Long Short-Term Memory (LSTM), and Lasso regression. This discovery highlights the notable effectiveness of the Random Forest algorithm in the specific context of Rajasthan for estimating crop yields, providing valuable guidance for agricultural decision-making and resource distribution.

D. The machine learning method mentioned in reference [4] looks at a variety of land-related and environmental factors to determine which crop would be best in a given area. The assessment of atmospheric conditions, including temperature, humidity, rainfall, and other meteorological variables, is part of this methodology. It also takes into account characteristics unique to the land, such as the pH and type of the soil as well as past records of crops that were grown there. The method seeks to offer suggestions to farmers by combining this varied collection of data, assisting them in choosing

crops that will maximize agricultural output in the particular area.

E. In reference [5] it mainly describes that the The farmer can utilize the prediction to determine the crop's yield prior to planting in an agricultural field. Future crop yield is predicted with accuracy using Random Forest, one of the most potent and well-liked supervised machine learning algorithms.

F. in reference [6], the authors aim to emphasize the significant contribution of the agricultural sector and its associated industries to the Indian economy. The success of crop cultivation hinges on numerous factors, including crop rotation, soil moisture levels, crop composition (including nutrients like nitrogen, phosphorus, and potassium), as well as weather conditions such as temperature, precipitation, and atmospheric elements. Modern agriculture leverages cutting-edge technologies like machine learning (ML) algorithms such as Random Forest, Decision Tree, and Artificial Neural Network to determine the most suitable crop choices for specific geographical regions.In order to improve predictive models, deep learning techniques have also been incorporated. These models not only predict crop yields but also offer important details regarding the amounts of soil nutrients needed and the costs involved.

G. Through the provision of cutting-edge tools for crop yield prediction prior to planting, the project aims to empower farmers. By developing the first version of an active prediction system, this prediction system aims to address issues in agriculture. Accurate crop yield predictions will be made by this system using machine learning algorithms and an intuitive graphical user interface that will be accessible via the web. Farmers will have easy access to the outcomes of these forecasts, empowering them to make well-informed decisions regarding their farming operations.

Many strategies and algorithms are used in the fields of data analytics and crop prediction. The algorithm known as Random Forest is a well-liked option due to its precision in forecasting crop yields, and it is one of the algorithms mentioned. It considers a range of elements, including

H. In the reference [8], the goal of the paper is to determine the best crop prediction model in order to assist farmers in selecting crops based on informed decisions. The study uses two different criteria, entropy and gini, to evaluate a variety of algorithms for machine learning, including Random Forest Classifier, Decision Tree, and K-Nearest Neighbor (KNN). The study's findings indicate that, of these algorithms, the Random Forest Classifier forecasts crop results with the highest degree of accuracy. This suggests that Random Forest is a promising choice for crop prediction, enabling farmers to optimize their crop choices based on factors like climatic conditions and soil nutrients.

III. METHODOLOGY

The primary objective of this system is to forecast crop yields within a particular geographic region, enabling farmers to make informed decisions regarding the most suitable crops to cultivate in their area. This methodology encompasses a comprehensive process that involves the utilization of data, technology, and predictive algorithms to estimate potential crop outcomes. The methodology includes:

A. Data Collection:

Gather historical information on crop yields. This information should include elements like the year, typical rainfall, use of pesticides, typical temperature, region (area), and the particular crop item. The dataset 'yield_df.csv' is loaded using Pandas.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot  as plt
import seaborn as sns
```

```
df = pd.read_csv('yield_df.csv')
```

B. Data Preprocessing:

i) Data Cleaning: Unwanted columns are dropped

```
df.drop('Unnamed: 0',axis=1,inplace=True)
```

ii) Handling Missing Values: Checking and handling any missing values that may exist.

```
df.isnull().sum()
```

iii) Data Type Conversion: Ensure that the 'average_rain_fall_mm_per_year' column has the appropriate data type (float64).

C. Exploratory Data Analysis (EDA): Visualizing data using Seaborn and Matplotlib to gain insights into the dataset.

```
plt.figure(figsize = (10,20))
sns.countplot(y=df['Area'])
```

```
country = (df['Area'].unique())
```

```
yield_per_country =[]
for state in country:
  yield_per_country .append(df[df['Area']==state]['hg/ha_yield'].sum())
```

```
yield_per_country
```

```
sns.countplot(y=df['Item'])
```

D. Feature Selection: Selecting relevant features and creating a new DataFrame 'df' with the chosen columns.

```
col = ['Year','average_rain_fall_mm_per_year','pesticides_tonnes','avg_temp','Area','Item','hg/ha_yield']
```

```
df = df[col]
```

```
df
```

E. Data splitting: Splitting the dataset into training and testing sets.

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
```

F. Data preprocessing (Scaling and Encoding): Use standardization and one-hot encoding to preprocess the features.

```
from sklearn.preprocessing import OneHotEncoder,StandardScaler
from sklearn.compose import ColumnTransformer
```

```
ohe = OneHotEncoder(drop = 'first')
scaler = StandardScaler()
```

```
preprocessor = ColumnTransformer(
    transformers =[('onehotencoder',ohe,[4,5]),
                   ('standrization',scaler,[0,1,2,3])
    ],
    remainder='passthrough'
    )
```

```
x_train_dummy = preprocessor.fit_transform(x_train)
x_test_dummy = preprocessor.transform(x_test)
```

G. Model Building: Train a Decision Tree Regressor model.

```
from sklearn.tree import DecisionTreeRegressor
```

```
'dtr' : DecisionTreeRegressor()
```

```
dtr = DecisionTreeRegressor()
dtr.fit(x_train_dummy,y_train)
dtr.predict(x_test_dummy)
```

H. Prediction Function: Create a function to make crop yield predictions.

```
def prediction(Year,average_rain_fall_mm_per_year,pesticides_tonnes,avg_temp,Area,Item):
    features = np.array([[Year,average_rain_fall_mm_per_year,pesticides_tonnes,avg_temp,Area,Item
    transformed_features = preprocessor.transform(features)
    predicted_value = dtr.predict(transformed_features).reshape(1,-1)
    return predicted_value[0]
```

```
Year = 2000
average_rain_fall_mm_per_year = 59.0
pesticides_tonnes = 3024.11
avg_temp = 26.55
Area = 'Saudi Arabia'
Item = 'Sorghum'

result = prediction(Year,average_rain_fall_mm_per_year,pesticides_tonnes,avg_temp,Area,It
```

```
result
```

I. Results

Figure1. Output User Interface

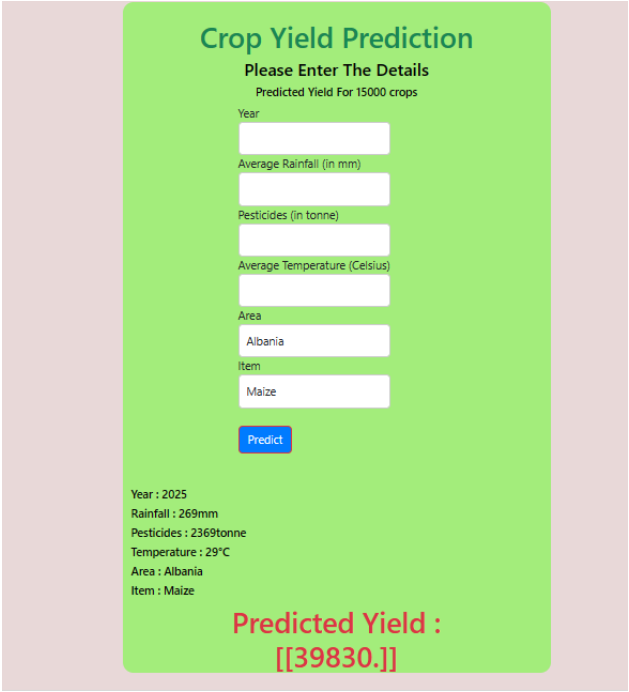

Figure2.Output Result

IV. CONCLUSION

crop yield prediction is a valuable application of machine learning. Through this system the farmers can easily identify which crop will produce the best yield by entering the features such as the climatic conditions, crop name, place.

Crop yield prediction using decision trees involves analyzing various factors such as weather conditions, pesticides etc. to make accurate predictions about the expected crop yield. Depending on the dataset, features, and particular decision tree algorithm used, the accuracy of crop yield prediction using decision tree models can vary. Based on our dataset the accuracy is 0.9769. The Decision Tree algorithm, in fact, can be used to forecast crop yields for a given crop in a given year by considering a number of features, such as pesticide use, climatic conditions, and other pertinent factors. Using this method makes it possible to create a predictive model that can be useful for agricultural decision-making and aid in the estimation of crop outcomes.

V. REFERENCES

[1]. Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber "Crop Yield Prediction Based on Indian Agriculture Using Machine Learning"2020 International Conference for Emerging Technology (INCET)

[2]. Suresh Kumar Sharma, Durga Prasad Sharma, Jitendra Kumar Verma "Study on Machine Learning Algorithms in Crop Yield Predictions specific to Indian Agricultural Contexts"2021 International Conference on Computational Performance Evaluvation (ComPE)

[3]. Kavita Jhajhariaa, Pratistha Mathura, Sanchit Jaina, Sukriti Nijhawan "Crop Yield Prediction Using Machine Learning and Deep Learning Techniques" International Conference on Machine Learning and Data Engineering

[4]. Sangeeta, Shruthi G "Design and Implementation of Crop Yield Prediction Model in Agriculture"

International Journal of Scientific and Technology Research Volume 8, Issue 01, January 2020.

[5]. Priya P, U Muthaiah, M Balamurugan “Predicting Yield of the Crop Yield Prediction Using Machine Learning Algorithm” International Journal of Engineering Science and Research Technology

[6]. Sonal Agarwal and Sandhya Tarar” a hybrid approach for crop yield prediction using machine learning and deep learning algorithms” Journal of Physics: Conference Series

[7]. Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, Mansing Rathod “Crop Yield Prediction Using Machine Learning” International Journal of Science and Research (IJSR)

[8]. Madhuri Shripathi Rao, Arushi Singh, N.V. Subba Reddy and Dinesh U Acharya” Crop prediction using machine learning” Journal of Physics: Conference Series