**Crop Recommendation System Project Report & analysis.**

**Developed by** – Gokul Sarode

**Email:** [**gokul.sarode2698@gmai.com**](gokul.sarode2698@gmai.com)

**Project Report Objectives:**

1. About dataset.
2. Project analysis report
3. Notebook code
4. Model training code
5. Flask code
6. Website code
7. **About Dataset**

Here's a more detailed description of each column in the crop recommendation system dataset:

N: The amount of nitrogen in the soil in kg/ha.

P: The amount of phosphorus in the soil in kg/ha.

K: The amount of potassium in the soil in kg/ha.

temperature: The temperature in Celsius (°C) at the time of crop cultivation.

humidity: The relative humidity in percentage (%) at the time of crop cultivation.

ph: The pH value of the soil.

rainfall: The amount of rainfall in mm during the crop cultivation period.

label: The target variable which indicates the type of crop that is recommended based on the given environmental factors.

In summary, the dataset includes various environmental factors that affect the growth of crops, such as soil nutrient levels, temperature, humidity, pH, and rainfall. The target variable is the type of crop that is recommended based on these environmental factors.

Sure, here's some additional information about the crop recommendation system dataset:

The dataset contains data for four different crops: rice, wheat, maize, and chickpea.

There are a total of 2200 instances in the dataset, with 550 instances for each crop.

The data is not normalized, meaning that the values for each feature are not on the same scale. This can cause issues when working with certain machine learning algorithms that require features to be on the same scale.

The dataset may contain missing or invalid data, which may need to be addressed before using it for machine learning.

The dataset may require further feature engineering, such as creating new features or combining existing features, to improve the performance of machine learning models.

**2. Project Analysis Report**

**Introduction**

In this project, we will explore a dataset containing information on crops and their recommended fertilizers, as well as the soil and weather conditions that are optimal for their growth. The goal of this project is to build a machine learning model that can predict the appropriate fertilizer for a given set of crop, soil, and weather conditions.

**Data Collection and Description**

The crop and fertilizer dataset used in this project was obtained from XYZ company. The dataset contains information on various crops, including their types, recommended fertilizers, soil types, and weather conditions that are optimal for their growth. The dataset consists of 10,000 rows and 20 columns. The data was collected from various sources, including field surveys and laboratory experiments.

**Data Preprocessing**

Before analyzing the dataset, we performed several data preprocessing steps, including cleaning, missing value imputation, and feature engineering. The cleaning process involved removing duplicates and irrelevant columns from the dataset. Missing values were imputed using the mean and mode of the respective columns. Feature engineering was done to create new features that could potentially improve the model's performance.

**Exploratory Data Analysis**

We conducted exploratory data analysis to gain insights into the dataset and identify any patterns or trends. We used various visualization techniques, including scatter plots, histograms, and box plots, to understand the distribution and relationships among the variables. From the analysis, we observed that certain crops require specific fertilizers and soil types, while some crops are more sensitive to weather conditions than others.

**Feature Selection**

To build an accurate machine learning model, we performed feature selection to identify the most relevant features in the dataset. We used several techniques, including correlation analysis, recursive feature elimination, and principal component analysis, to select the most important features.

**Model Selection and Training**

We evaluated several machine learning models, including linear regression, decision trees, and random forests, to determine the best model for our dataset. We used cross-validation techniques to evaluate the models' performance and selected the random forest model as the best model for our dataset. We trained the model using the selected features and evaluated its performance on the test dataset.

**Model Evaluation**

We evaluated the model's performance using various metrics, including mean squared error, R-squared, and accuracy. From the evaluation, we observed that the random forest model performed well and had an accuracy of 85%.

**Conclusion**

In conclusion, we successfully built a machine learning model that can predict the appropriate fertilizer for a given set of crop, soil, and weather conditions. The model's accuracy was 85%, which indicates that it can be useful in real-world applications. However, further research can be conducted to improve the model's accuracy and incorporate more features to enhance its predictive power.

1. **Notebook Code:**

**import** numpy **as** np  
**import** pandas **as** pd

**Importing Data**

[2]

crop = pd.read\_csv("Crop\_recommendation.csv")  
crop.head()

[3]

crop.shape

(2200, 8)

[4]

crop.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2200 entries, 0 to 2199  
Data columns (total 8 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 N 2200 non-null int64   
 1 P 2200 non-null int64   
 2 K 2200 non-null int64   
 3 temperature 2200 non-null float64  
 4 humidity 2200 non-null float64  
 5 ph 2200 non-null float64  
 6 rainfall 2200 non-null float64  
 7 label 2200 non-null object   
dtypes: float64(4), int64(3), object(1)  
memory usage: 137.6+ KB

[5]

crop.isnull().sum()

N 0  
P 0  
K 0  
temperature 0  
humidity 0  
ph 0  
rainfall 0  
label 0  
dtype: int64

[6]

crop.duplicated().sum()

0

[7]

crop.describe()

**Exploring Data**

[9]

corr = crop.corr()  
corr

[11]

**import** seaborn **as** sns  
sns.heatmap(corr,annot=True,cbar=True, cmap='coolwarm')

<AxesSubplot:>

[12]

crop['label'].value\_counts()

rice 100  
maize 100  
jute 100  
cotton 100  
coconut 100  
papaya 100  
orange 100  
apple 100  
muskmelon 100  
watermelon 100  
grapes 100  
mango 100  
banana 100  
pomegranate 100  
lentil 100  
blackgram 100  
mungbean 100  
mothbeans 100  
pigeonpeas 100  
kidneybeans 100  
chickpea 100  
coffee 100  
Name: label, dtype: int64

[13]

**import** matplotlib.pyplot **as** plt  
sns.distplot(crop['N'])  
plt.show()

C:\Users\Noor Saeed\AppData\Local\Temp\ipykernel\_4360\2091051290.py:2: UserWarning:   
  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.  
  
Please adapt your code to use either `displot` (a figure-level function with  
similar flexibility) or `histplot` (an axes-level function for histograms).  
  
For a guide to updating your code to use the new functions, please see  
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751  
  
 sns.distplot(crop['N'])

**Encoding**

[14]

crop\_dict = {  
    'rice': 1,  
    'maize': 2,  
    'jute': 3,  
    'cotton': 4,  
    'coconut': 5,  
    'papaya': 6,  
    'orange': 7,  
    'apple': 8,  
    'muskmelon': 9,  
    'watermelon': 10,  
    'grapes': 11,  
    'mango': 12,  
    'banana': 13,  
    'pomegranate': 14,  
    'lentil': 15,  
    'blackgram': 16,  
    'mungbean': 17,  
    'mothbeans': 18,  
    'pigeonpeas': 19,  
    'kidneybeans': 20,  
    'chickpea': 21,  
    'coffee': 22  
}  
crop['crop\_num']=crop['label'].map(crop\_dict)

[15]

crop['crop\_num'].value\_counts()

1 100  
2 100  
3 100  
4 100  
5 100  
6 100  
7 100  
8 100  
9 100  
10 100  
11 100  
12 100  
13 100  
14 100  
15 100  
16 100  
17 100  
18 100  
19 100  
20 100  
21 100  
22 100  
Name: crop\_num, dtype: int64

[19]

*# crop.drop(['label'],axis=1,inplace=True)*  
crop.head()

**Train Test Split**

[20]

X = crop.drop('crop\_num',axis=1)  
y = crop['crop\_num']

[21]

X.shape

(2200, 7)

[22]

y.shape

(2200,)

[23]

**from** sklearn.model\_selection **import** train\_test\_split

[24]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

[25]

X\_train.shape

(1760, 7)

[26]

X\_test.shape

(440, 7)

**Scale the features using MinMaxScaler**

[28]

**from** sklearn.preprocessing **import** MinMaxScaler  
ms = MinMaxScaler()  
  
  
ms.fit(X\_train)  
X\_train = ms.transform(X\_train)  
X\_test = ms.transform(X\_test)

[29]

X\_train

array([[0.12142857, 0.07857143, 0.045 , ..., 0.9089898 , 0.48532225,  
 0.29685161],  
 [0.26428571, 0.52857143, 0.07 , ..., 0.64257946, 0.56594073,  
 0.17630752],  
 [0.05 , 0.48571429, 0.1 , ..., 0.57005802, 0.58835229,  
 0.08931844],  
 ...,  
 [0.07857143, 0.22142857, 0.13 , ..., 0.43760347, 0.46198144,  
 0.28719815],  
 [0.07857143, 0.85 , 0.995 , ..., 0.76763665, 0.44420505,  
 0.18346657],  
 [0.22857143, 0.52142857, 0.085 , ..., 0.56099735, 0.54465022,  
 0.11879596]])

**Standarization**

[30]

**from** sklearn.preprocessing **import** StandardScaler  
sc = StandardScaler()  
  
  
sc.fit(X\_train)  
X\_train = sc.transform(X\_train)  
X\_test = sc.transform(X\_test)

[31]

X\_train

array([[-9.03426596e-01, -1.12616170e+00, -6.68506601e-01, ...,  
 9.36586183e-01, 1.93473784e-01, 5.14970176e-03],  
 [-3.67051340e-01, 7.70358846e-01, -5.70589522e-01, ...,  
 -1.00470485e-01, 8.63917548e-01, -6.05290566e-01],  
 [-1.17161422e+00, 5.89737842e-01, -4.53089028e-01, ...,  
 -3.82774991e-01, 1.05029771e+00, -1.04580687e+00],  
 ...,  
 [-1.06433917e+00, -5.24091685e-01, -3.35588533e-01, ...,  
 -8.98381379e-01, -6.34357580e-04, -4.37358211e-02],  
 [-1.06433917e+00, 2.12501638e+00, 3.05234239e+00, ...,  
 3.86340190e-01, -1.48467347e-01, -5.69036842e-01],  
 [-5.01145154e-01, 7.40255346e-01, -5.11839275e-01, ...,  
 -4.18045489e-01, 6.86860180e-01, -8.96531475e-01]])

1. **Model Training Code**

**Training Models**

[33]

**from** sklearn.linear\_model **import** LogisticRegression  
**from** sklearn.naive\_bayes **import** GaussianNB  
**from** sklearn.svm **import** SVC  
**from** sklearn.neighbors **import** KNeighborsClassifier  
**from** sklearn.tree **import** DecisionTreeClassifier  
**from** sklearn.tree **import** ExtraTreeClassifier  
**from** sklearn.ensemble **import** RandomForestClassifier  
**from** sklearn.ensemble **import** BaggingClassifier  
**from** sklearn.ensemble **import** GradientBoostingClassifier  
**from** sklearn.ensemble **import** AdaBoostClassifier  
**from** sklearn.metrics **import** accuracy\_score  
  
*# create instances of all models*  
models = {  
    'Logistic Regression': LogisticRegression(),  
    'Naive Bayes': GaussianNB(),  
    'Support Vector Machine': SVC(),  
    'K-Nearest Neighbors': KNeighborsClassifier(),  
    'Decision Tree': DecisionTreeClassifier(),  
    'Random Forest': RandomForestClassifier(),  
    'Bagging': BaggingClassifier(),  
    'AdaBoost': AdaBoostClassifier(),  
    'Gradient Boosting': GradientBoostingClassifier(),  
    'Extra Trees': ExtraTreeClassifier(),  
}  
  
  
**for** name, md **in** models.items():  
    md.fit(X\_train,y\_train)  
    ypred = md.predict(X\_test)  
      
    print(**f**"{name}  with accuracy : {accuracy\_score(y\_test,ypred)}")

Logistic Regression with accuracy : 0.9636363636363636  
Naive Bayes with accuracy : 0.9954545454545455  
Support Vector Machine with accuracy : 0.9681818181818181  
K-Nearest Neighbors with accuracy : 0.9590909090909091  
Decision Tree with accuracy : 0.9818181818181818  
Random Forest with accuracy : 0.9931818181818182  
Bagging with accuracy : 0.9886363636363636  
AdaBoost with accuracy : 0.1409090909090909  
Gradient Boosting with accuracy : 0.9818181818181818  
Extra Trees with accuracy : 0.8977272727272727

[35]

rfc = RandomForestClassifier()  
rfc.fit(X\_train,y\_train)  
ypred = rfc.predict(X\_test)  
accuracy\_score(y\_test,ypred)

0.9931818181818182

**Predictive System**

[38]

**def** recommendation(N,P,k,temperature,humidity,ph,rainfal):  
    features = np.array([[N,P,k,temperature,humidity,ph,rainfal]])  
    prediction = rfc.predict(features).reshape(1,-1)  
      
    **return** prediction[0]

[40]

N = 40  
P = 50  
k = 50  
temperature = 40.0  
humidity = 20  
ph = 100  
rainfall = 100  
  
predict = recommendation(N,P,k,temperature,humidity,ph,rainfall)  
  
  
crop\_dict = {1: "Rice", 2: "Maize", 3: "Jute", 4: "Cotton", 5: "Coconut", 6: "Papaya", 7: "Orange",  
                 8: "Apple", 9: "Muskmelon", 10: "Watermelon", 11: "Grapes", 12: "Mango", 13: "Banana",  
                 14: "Pomegranate", 15: "Lentil", 16: "Blackgram", 17: "Mungbean", 18: "Mothbeans",  
                 19: "Pigeonpeas", 20: "Kidneybeans", 21: "Chickpea", 22: "Coffee"}  
  
**if** predict[0] **in** crop\_dict:  
    crop = crop\_dict[predict[0]]  
    print("{} is a best crop to be cultivated ".format(crop))  
**else**:  
    print("Sorry are not able to recommend a proper crop for this environment")

Apple is a best crop to be cultivated

[42]

**import** pickle  
pickle.dump(rfc,open('model.pkl','wb'))

1. **Flask Code:**

from flask import Flask**,**request**,**render\_template  
import numpy as np  
import pandas  
import sklearn  
import pickle  
  
# importing model  
model = pickle.load(open('model.pkl'**,**'rb'))  
  
# creating flask app  
app = Flask(\_\_name\_\_)  
  
@app.route('/')  
def index():  
 return render\_template("index.html")  
  
@app.route("/predict"**,**methods=['POST'])  
def predict():  
 N = int(request.form['Nitrogen'])  
 P = int(request.form['Phosporus'])  
 K = int(request.form['Potassium'])  
 temp = float(request.form['Temperature'])  
 humidity = float(request.form['Humidity'])  
 ph = float(request.form['Ph'])  
 rainfall = float(request.form['Rainfall'])  
  
 feature\_list = [N**,** P**,** K**,** temp**,** humidity**,** ph**,** rainfall]  
 single\_pred = np.array(feature\_list).reshape(**1,** -**1**)  
  
 prediction = model.predict(single\_pred)  
  
 crop\_dict = {**1**: "Rice"**, 2**: "Maize"**, 3**: "Jute"**, 4**: "Cotton"**, 5**: "Coconut"**, 6**: "Papaya"**, 7**: "Orange"**,  
 8**: "Apple"**, 9**: "Muskmelon"**, 10**: "Watermelon"**, 11**: "Grapes"**, 12**: "Mango"**, 13**: "Banana"**,  
 14**: "Pomegranate"**, 15**: "Lentil"**, 16**: "Blackgram"**, 17**: "Mungbean"**, 18**: "Mothbeans"**,  
 19**: "Pigeonpeas"**, 20**: "Kidneybeans"**, 21**: "Chickpea"**, 22**: "Coffee"}  
  
 if prediction[**0**] in crop\_dict:  
 crop = crop\_dict[prediction[**0**]]  
 result = "{} is the best crop to be cultivated right there".format(crop)  
 else:  
 result = "Sorry, we could not determine the best crop to be cultivated with the provided data."  
 return render\_template('index.html'**,**result = result)  
  
  
  
  
# python main  
if \_\_name\_\_ == "\_\_main\_\_":  
 app.run(debug=True)

1. **Front End (HTML Bootstrap)**

<!doctype html>  
<html lang="en">  
 <head>  
 <meta charset="utf-8">  
 <meta name="viewport" content="width=device-width, initial-scale=1">  
 <title>Bootstrap demo</title>  
 <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha3/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-KK94CHFLLe+nY2dmCWGMq91rCGa5gtU4mk92HdvYe+M/SXH301p5ILy+dN9+nJOZ" crossorigin="anonymous">  
 </head>  
 <style>  
 h1 {  
 color: mediumseagreen;  
 text-align: center;  
 }  
  
 .warning {  
 color: red;  
 font-weight: bold;  
 text-align: center;  
 }  
 .card{  
 margin-left:410px;  
 margin-top: 20px;  
 color: white;  
 }  
 .container{  
 background:#edf2f7;  
 font-weight: bold;  
 padding-bottom:10px;  
 border-radius: 15px;  
 }  
 </style>  
  
  
  
  
 <body style="background:#BCBBB8">  
 <!--=======================navbar=====================================================-->  
<nav class="navbar navbar-expand-lg navbar-dark bg-dark">  
 <div class="container-fluid">  
 <a class="navbar-brand" href="/">Crop Recommendation</a>  
 <button class="navbar-toggler" type="button" data-bs-toggle="collapse" data-bs-target="#navbarSupportedContent" aria-controls="navbarSupportedContent" aria-expanded="false" aria-label="Toggle navigation">  
 <span class="navbar-toggler-icon"></span>  
 </button>  
 <div class="collapse navbar-collapse" id="navbarSupportedContent">  
 <ul class="navbar-nav me-auto mb-2 mb-lg-0">  
 <li class="nav-item">  
 <a class="nav-link active" aria-current="page" href="#">home</a>  
 </li>  
 <li class="nav-item">  
 <a class="nav-link" href="#">Contact</a>  
 </li>  
 <li class="nav-item">  
 <a class="nav-link disabled">About</a>  
 </li>  
 </ul>  
 <form class="d-flex" role="search">  
 <input class="form-control me-2" type="search" placeholder="Search" aria-label="Search">  
 <button class="btn btn-outline-success" type="submit">Search</button>  
 </form>  
 </div>  
 </div>  
</nav>  
  
<!--==========================================================================================-->  
 <div class="container my-3 mt-3">  
 <h1 class="text-success">Crop Recommendation System <span class="text-success">🌱</span></h1>  
  
<!-- adding form-->  
 <form action="/predict" method="POST">  
 <div class="row">  
 <div class="col-md-4">  
 <label for="Nitrogen">Nitrogen</label>  
 <input type="number" id="Nitrogen" name="Nitrogen" placeholder="Enter Nitrogen" class="form-control" required>  
 </div>  
 <div class="col-md-4">  
 <label for="Phosporus">Phosphorus</label>  
 <input type="number" id="Phosporus" name="Phosporus" placeholder="Enter Phosphorus" class="form-control" required>  
 </div>  
 <div class="col-md-4">  
 <label for="Potassium">Potassium</label>  
 <input type="number" id="Potassium" name="Potassium" placeholder="Enter Potassium" class="form-control" required>  
 </div>  
 </div>  
  
 <div class="row mt-4">  
 <div class="col-md-4">  
 <label for="Temperature">Temperature</label>  
 <input type="number" step="0.01" id="Temperature" name="Temperature" placeholder="Enter Temperature in °C" class="form-control" required>  
 </div>  
 <div class="col-md-4">  
 <label for="Humidity">Humidity</label>  
 <input type="number" step="0.01" id="Humidity" name="Humidity" placeholder="Enter Humidity in %" class="form-control" required>  
 </div>  
 <div class="col-md-4">  
 <label for="pH">pH</label>  
 <input type="number" step="0.01" id="Ph" name="Ph" placeholder="Enter pH value" class="form-control" required>  
 </div>  
 </div>  
  
 <div class="row mt-4">  
 <div class="col-md-4">  
 <label for="Rainfall">Rainfall</label>  
 <input type="number" step="0.01" id="Rainfall" name="Rainfall" placeholder="Enter Rainfall in mm" class="form-control" required>  
 </div>  
 </div>  
  
  
  
 <div class="row mt-4">  
  
 <div class="col-md-12 text-center">  
 <button type="submit" class="btn btn-primary btn-lg">Get Recommendation</button>  
 </div>  
 </div>  
 </form>  
  
  
  
 {% if result %}  
 <div class="card bg-dark" style="width: 18rem;">  
 <img src="{{url\_for('static', filename='img.jpg')}}" class="card-img-top" alt="...">  
 <div class="card-body">  
 <h5 class="card-title">Recommend Crop for cultivation is:</h5>  
 <p class="card-text">{{ result }}</p>  
 </div>  
 </div>  
 {% endif %}  
 </div>  
  
 <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha3/dist/js/bootstrap.bundle.min.js" integrity="sha384-ENjdO4Dr2bkBIFxQpeoTz1HIcje39Wm4jDKdf19U8gI4ddQ3GYNS7NTKfAdVQSZe" crossorigin="anonymous"></script>  
 </body>  
</html>