CMPE - 256 Advanced Data Mining

Crop Recommendation System

Project Report submitted by

DATA FARMERS

Devi Priya Ravi
Dhanasree Rajamani
Sravani Thota

Demo URL

https://drive.google.com/file/d/1V1vN0yQPV5ErdW_oJSw9j_W6YVM0L1Aq/view?usp=share_link

Introduction

Agriculture is considered as an important pillar of the global economy, meeting one of the basic human needs along with food. In most countries it is considered the most important source of employment. Many countries such as India still use traditional farming methods, and farmers don't rely on advanced techniques during farming due to lack of knowledge, high costs, or ignorance of the benefits of these techniques. Lack of knowledge about soil types, yields, crops, weather, improper use of pesticides, irrigation problems, harvest failures and lack of information about market developments lead farmers to losses and incurred additional costs. Lack of knowledge at each stage of agriculture creates new problems or amplifies old ones, increasing the cost of agriculture. Daily population growth is also increasing pressure on agriculture. Precision learning in agriculture is very important for improving the overall yield of crops.

Business Understanding

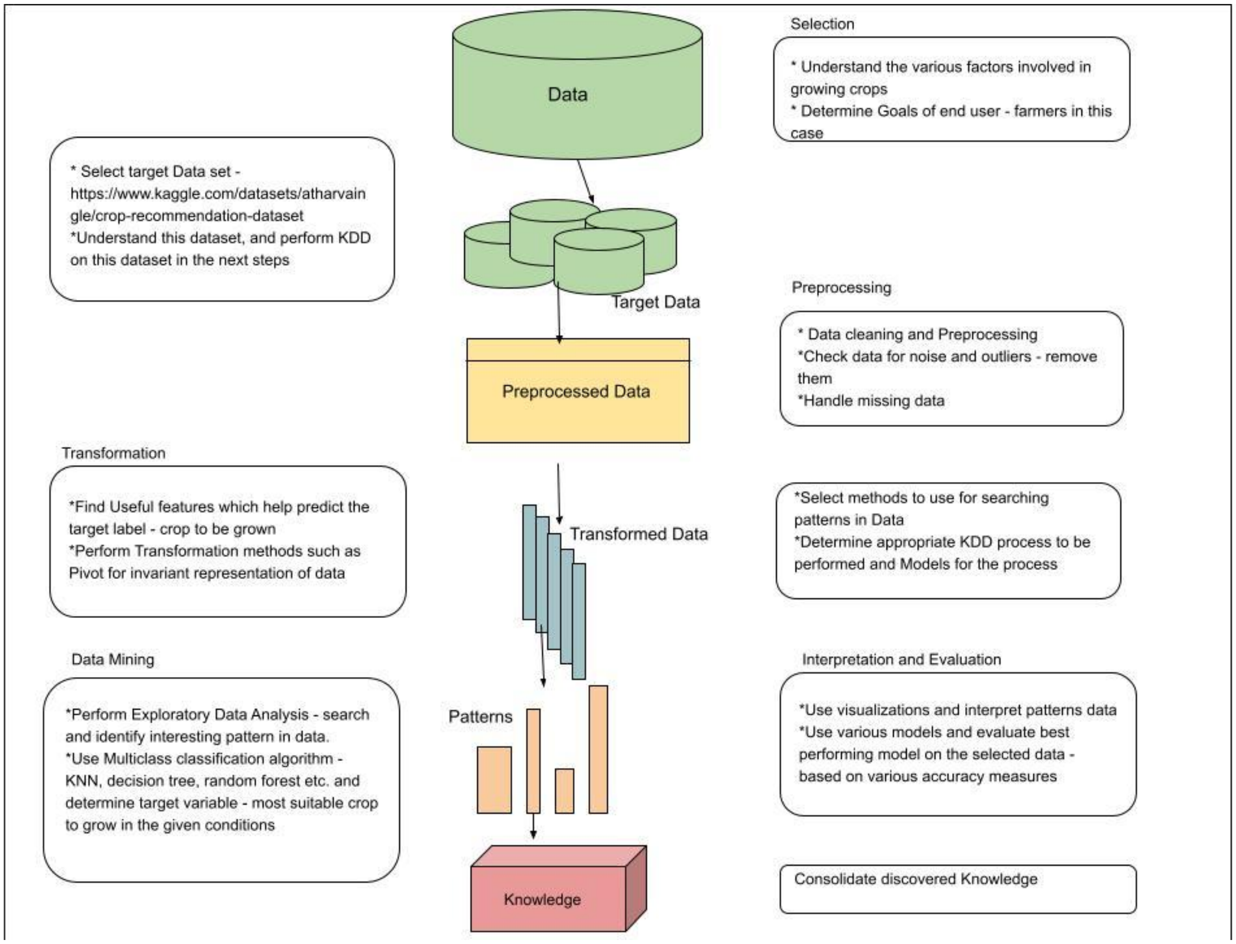
This project aims to incorporate science in farming - Crop Recommendation system - which uses various data features to determine the most suitable crop to grow in a given environment. This helps farmers make informed decisions about farming strategies.

Colab

https://colab.research.google.com/drive/1oCAiaB0ugTSWgfA8oNzdVEejjVOD9nNN#scrollTo=jGF8SIJnq7nw&uniqifier=1

Architecture

Knowledge Discovery Steps performed in the project



Data

The dataset consists of the following parameters:

N: Nitrogen content in soil
P: Phosphorus content in soil
K: Potassium content in soil
Temperature: Temperature in degree Celsius
Humidity: Relative humidity in percentage
pH: Indicate the chemical composition of soil

Rainfall: Rainfall in mm

Target label denotes the most suitable crop that can be grown in the given environment.

https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

Data Understanding

We have performed various steps such as data preparation, exploratory data analysis and feature engineering.

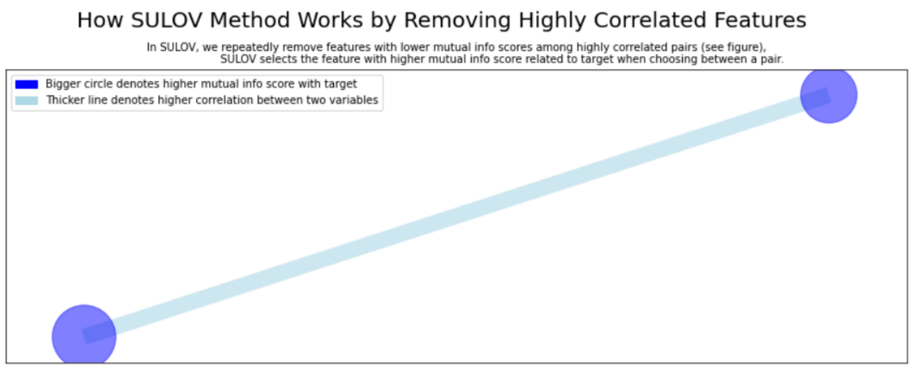
Data Science Algorithms

We have used the following algorithms :

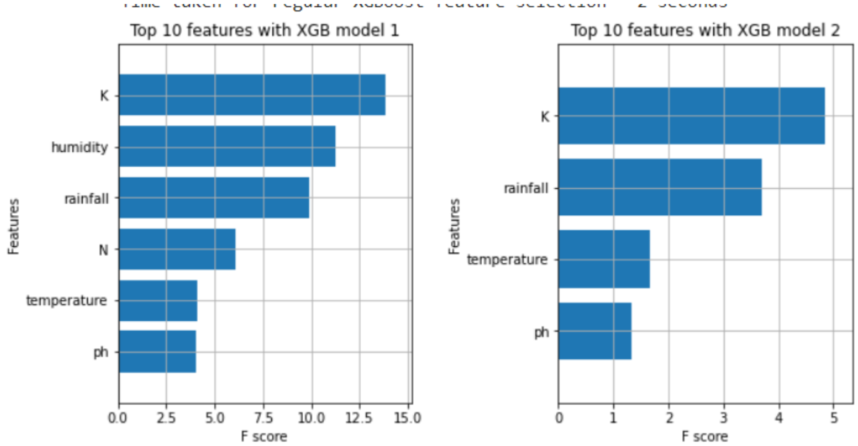
- Decision Tree
- Gaussian Naive Bayes
- Support Vector Machine
- Logistic Regression
- Random Forest

Features Used

Feature engineering is performed to determine the important features from the dataset, and understand how individual features in the dataset influence the target label. FeatureWiz has been used to perform feature engineering.




```
Time taken for SULOV method = 0 seconds
    Adding 0 categorical variables to reduced numeric variables  of 6
Final list of selected 6 vars after SULOV = ['N', 'humidity', 'ph', 'rainfall', 'temperature', 'K']
Converting all features to numeric before sending to XGBoost...
```




```
    Completed XGBoost feature selection in 1 seconds
#######################################################################################
#####          F E A T U R E   S E L E C T I O N   C O M P L E T E D          #######
#######################################################################################
Selected 3 important features:
['K', 'humidity', 'rainfall']
Total Time taken for featurewiz selection = 5 seconds
Output contains a list of 3 important features and a train dataframe
```

Model deployment

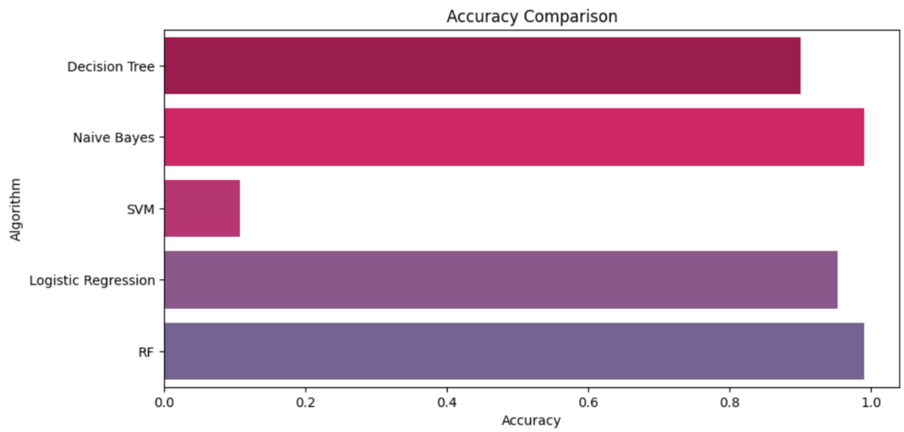
Various models are used to train the dataset and the model that works the best on the given dataset is determined based on accuracy score, precision, recall, f-score and support. Random forest classifier is determined as the best model for the dataset. The Random forest classifier model is saved and used for future predictions.

- **Random Forest**

[81]

```python
from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(Xtrain,Ytrain)

predicted_values = RF.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x)

print(classification_report(Ytest,predicted_values))
```

```
RF's Accuracy is:  0.990909090909091
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       0.94      1.00      0.97        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      1.00      1.00        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       0.90      1.00      0.95        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       1.00      1.00      1.00        23
       maize       1.00      1.00      1.00        21
       mango       1.00      1.00      1.00        26
```



Save The Model

Now Lets Save the model. We cannot save the model directly. Instead we need to convert into '.bin' format using pickle.
Pickle - A built-in library to save python objects.

```
[ ]  import pickle
```

```
[ ]  C = '1.0'
```

```
[ ]  output_file = f'rf_model_C={C}.bin'
     output_file
```

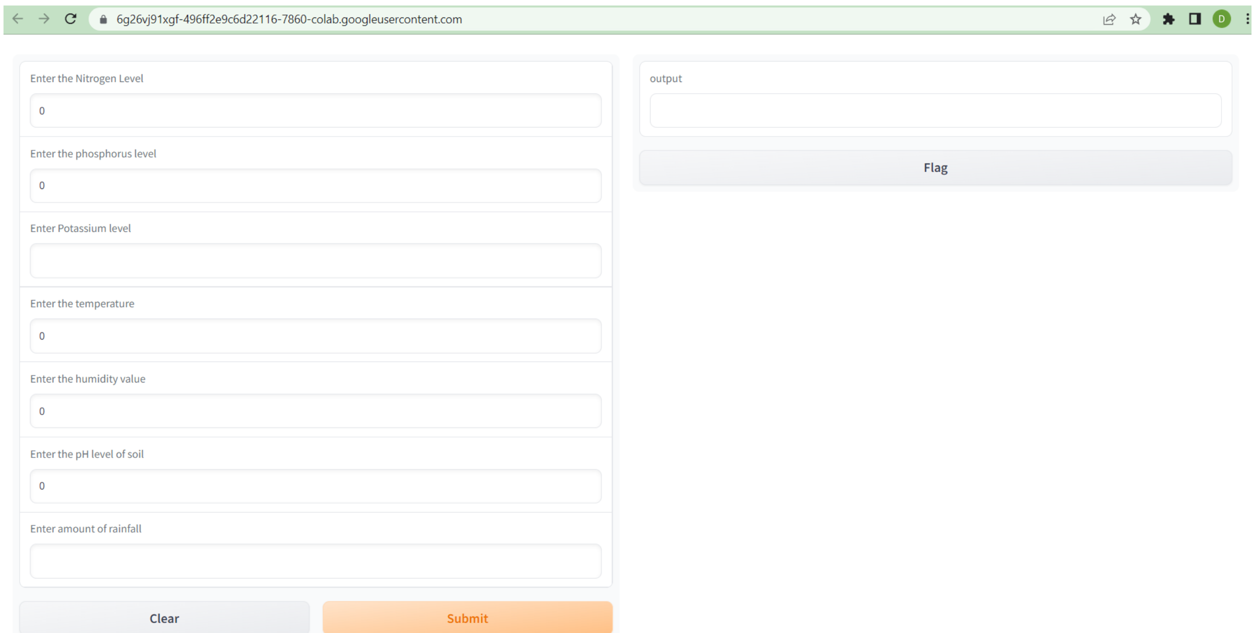
```
'rf_model_C=1.0.bin'
```

```
[ ]  with open(output_file, 'wb') as f_out:
       pickle.dump(RF, f_out)
```

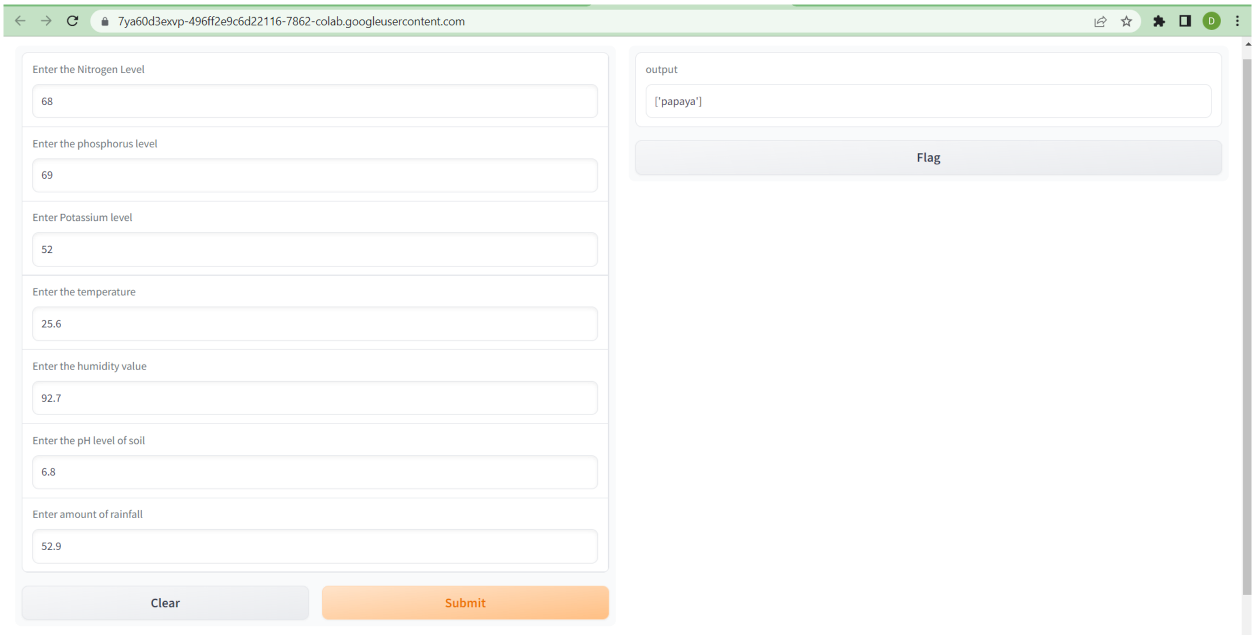
Client Side Design

Gradio app has been used to obtain input from users and predict the suitable crop to grow based on the given input environmental conditions.

Gradio app UI



Input values given and obtaining prediction in the Gradio app



Github

The Github link contains:

- Colab of the project
- Demo Video URL

https://github.com/Dhanasree-Rajamani/AdvancedDataMining/tree/main/256%20Crop%20Recommendation%20System

KDD and Feature Engineering

https://docs.google.com/document/d/1qAmPHvwAOOkIQq-H4DXDt1MWOpdeJALtu_Kyj9aCL64/edit?usp=sharing

Architecture, Component level design, Data flow diagram

https://docs.google.com/document/d/1sNV47wQEdmaKTivL6dt7asGigvQ7AlsAbxyh3YNX6wA/edit?usp=sharing