TITLE

SEMINAR REPORT

Submitted by

B. Hulda [RA1811003020542]
B. Pavan [RA1811003020543]

In partial fulfillment for the award of the degree of
BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE & ENGINEERING
of
FACULTY OF ENGINEERING AND TECHNOLOGY

S.R.M. Nagar, Kattankulathur, Kancheepuram District
MAY

S.R.M. Nagar, Kattankulathur, Kancheepuram District

MAY 2020

SRM INSTITUTE OF SCIENCE AND
TECHNOLOGY
(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that the Industrial training report titled " **Title: Crop prediction through soil classification** " is the bonafide work of " **B. Hulda [RA1811003020542] & B.Pavan [RA1811003020543]** " submitted for the course 18CSP103L Seminar – II. This report is a record of successful completion of the specified course evaluated based on literature reviews and the supervisor. No part of the Seminar Report has been submitted for a ny degree, diploma, title, or recognition before

SIGNATURE **SIGNATURE**

SUPERVISOR **HEAD OF THE DEPARTMENT INTERNAL EXAMINER I**
INTERNAL EXAMINER I I

TABLE OF CONTENTS :

Abstract

India is the land of agriculture and it is the major source of economy.70% of Indian population directly relies on agriculture. The common problem existing among the young Indian farmers is to choose the right crop based on the soil requirements. Due to This, they face a serious setback in productivity. Our work proposes to help farmers determine the soil quality by doing analysis on its various parameters and to suggest crops based on the results obtained using machine learning approach. In this we are using the machine learning algorithm such as Decision Tree Algorithm to improve the efficiency of Crop Suggestion System along with the powerful python library Scikit-Learn.

Problem Statement

Agriculture is the backbone of India. As we known, food stands first in the basic need for survival. Agriculture sector needs to be given the highest preference in development. There are several factors that affect agriculture. Some of them are soil, climate and land relief. Soil is a critical part of successful agriculture and is the original source of the nutrients that we use to grow crops. The nutrients move from the soil into plants that we eat like tomatoes. Farmers must use certain practices to make sure they are taking good care of the soil. The selection of right crop suitable for the fertile land must be ensured to make the best use of the practices followed.

Existing System

• Analysis of Crop prediction using soil classification.

• Varieties of Crop Data was analyzed for the Prediction of Crop and crop price.

• KNN algorithm was used for prediction.

• Data of different places was used for prediction.

• Data set used was taken from Govn.in

• The existing system predicts the crop yield by using the soil parameters and also Recommends fertiliser.

• It uses the crop yield information to make the end users Decide on the crop to be sown. Hence the system is not simple enough for dilettante Farmers to understand.

Advantages :

• It can be apply to the data from any distribution.

Example : data does not have to be separate with a linear boundary. very simple and intuitive.

• good classification if the no.of samples is large enough. provides good generalization accuracy on many domains. easy to understand which facilitates implementation and modification.

• K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

Disadvantages :

• Accuracy depends on the quality of the data.

• With large data, the prediction stage might be slow.

• Sensitive to the scale of the data and irrelevant features.

• Require high memory – need to store all of the training data.

• Given that it stores all of the training, it can be computationally expensive

Proposed System

• The proposed system obtains the soil and crop parameters and maps those to list The suitable crops. It passes the various inputs to the controller which uses the Decision Tree Algorithm for classification. The proposed system provides easy accessibility to the Users. They are also easy to use and under- stand by the dilettante farmers.

• Classification is a two-step process, learning step and prediction step, in machine learning.

• In the learning step, the model is developed based on given training data.

• In the prediction step, the model is used to predict the response for given data. Decision Tree is one of the easiest and popular classification algorithms to understand and interpret.

• Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

• The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).

In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

• Types of Decision Trees

• Types of decision trees are based on the type of target variable we have. It can be of two types:

1. **Categorical Variable Decision Tree:** Decision Tree which has a categorical target variable then it called a Categorical variable decision tree.

2. Continuous Variable Decision Tree:

Decision Tree has a continuous target variable then it is called Continuous Variable Decision Tree

Advantages of proposed System

• It overcomes the disadvantages of existing system.

• Compared to other algorithms decision trees requires less effort for data preparation during pre-processing.

• A decision tree does not require normalization of data.

• A decision tree does not require scaling of data as well. Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent.

• A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders.

Disadvantages of Proposed System

• 100% accuracy is not achieved in this system.

• A small change in the data can cause a large change in the structure of the decision tree causing instability.

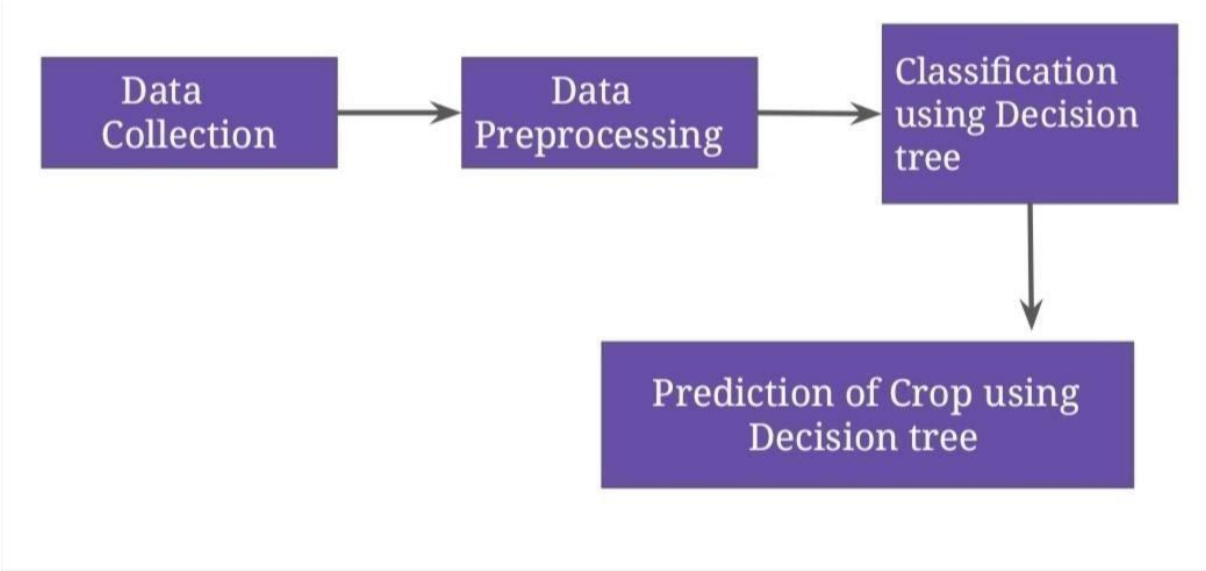
• For a Decision tree sometimes calculation can go far more complex compared to other algorithms.

• Decision tree often involves higher time to train the model.

• Decision tree training is relatively expensive as the complexity and time has taken are more.

• The Decision Tree algorithm is inadequate for applying regression and predicting continuous values.

Architecture Diagram



LITERATURE SURVEY

TITLE - Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique.

AUTHOR - Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr. N K Cauvery. In 2018

DESCRIPTION - In this paper, authors have mainly focussed on analysing the soil dataset that comprise of physical and chemical characteristic of soil and also various climatic condition to build a crop recommendation system using ensemble modelling and majority voting technique, to improve the crop productivity.

TITLE - Comparison of Machine Learning Algorithms for Soil Type Classification.

AUTHOR - Pramudyana Agus Harlianto, Noor Akhmad Setiawan Teguh Bharata Adji, In 2017.

DESCRIPTION - In this paper, they conducted a study of machine learning algorithms such as neural network, decision tree, naïve bayes, and SVM could be used to automate soil type classification with satisfactory accuracy (> 70%). using standard algorithm in RapidMiner, SVM is the best performance for classifying soil type. Attribute selection did not improve SVM"s accuracy as well as class reduction did not improve significantly.

TITLE - Machine Learning in Soil Classification and Crop Detection

AUTHOR - AshwiniRao, Janhavi.U, AbhishekGowda.N.S, Manjunatha, Mrs.Rafega Beham.A

DESCRIPTION -They developed a grading model for classification of soil samples based on support vector machine (SVM) using distinct scientific features. Different algorithms and features and filters are used to obtain and process color images of the soil samples. Color, texture etc. are the different features extracted using these algorithms. (SVM) uses only some of the training samples which lies at the edges of the class distribution in feature space and fit an optimal hyperplane between the classes. The correctness of the supervised classification is dependent on the training data used.

TITLE - Machine Learning Approaches For Soil Classification In A Multi - Agent Deficit Irrigation Control System.

AUTHOR - Daniel Smith, Wei Peng, In 2020

DESCRIPTION - In this paper, They used Techniques like Naïve Bayes classifier and the Bayes network, sequential minimal optimization, svm. Advantage is in machine learning approach to classify the textural composition of soil within the field using only soil moisture observations. Limitations is it Could be used to provide soil moisture predictions for deficit irrigation forecasts.

TITLE - Automated Soil Classification And Identification Using Machine Vision.
AUTHOR - M van Rooyen, N Luwes, In 2020.
DESCRIPTION - In this paper, They used Techniques like Image acquisition setup, Original image of cylinder during the sedimentation process. They identified as a possible approach for the machine vision system. Here the limitations is that the room temperature are important factors that should be noted and replicated as closely as possible for more accurate results.

TITLE - Soil Type Classification and Mapping using Hyperspectral.

AUTHOR - Amol.D.Vibhute.K.V.Kale, Rajesh.K.Dhumal, S.C.Mehrotra. In 2015

DESCRIPTION - In this paper, they proposed a method for identification, mapping and classification of several types of soil using SVM. In this they found that, support vector machine algorithm is advantageous for high dimensional datasets with less number of training samples. The disadvantage in this is to rectify atmospheric error which usually exist in hypersion data and only a small dataset can be used.

TITLE - Comparative Analysis Of Comparative Techniques On Soil Data To Predict Fertility Rate.

AUTHOR - Vrushali Bhuyar, In 2014.

DESCRIPTION - In this paper, They conducted a study of classification of soil fertility rate using J48, Naïve Bayes, and Random forest algorithm. J48 algorithm gives mostly good result than other algorithms. J48 algorithm in decision tree helps the farmer and decision makers to identify the the fertility rate of soil and on the nutrients found in the soil sample different fertilizers can be recommended.

TITLE - Machine Learning in Soil Classification.

AUTHOR - B. Bhattacharya, D. P. Solomatime,In 2020.

DESCRIPTION - In this paper, They used DT, ANN and SVM algorithms they are effectively mimics experts' classification procedure and automates the classification task.
The measured parameters classification based on the measured parameters was not possible

TITLE - An Intelligent Modelfor Indian Soil Classification using Various Machine Learning Techniques.

AUTHOR - Chandan, Rituala Thakur,In 2018.

DESCRIPTION - He proposed a system for imageclassification of soil based on soil images. The initial step is to gather soil test pictures which is the first important step in soil classification based on image processing, because it needs to consider factors such as scale and characteristics of soil under study. features of each type of soil is collects and stored in a separate database. This database is later used in the final stage for soil classification availability of ground reference data, complexity of data being studied,spatial resolution of the collected images are to be considered .The drawback related to the system is, it takes long training time for large datasets and choosing a good kernel Function is not easy.

TITLE - Soil Moisture Retrieval using UWB Echoes via Fuzzy Logic and Machine Learning

AUTHOR - Jing Liang, Senior Member, IEEE, Xiaoxu Liu, Kuo Liao. In 2020.

DESCRIPTION - In this paper, They used Techniques like T1FLS, ANFIS(Adaptive Neuro Fuzzy Inference System) , FL(Fuzzy Logic) ANFIS's features are corrupted by noise, it still shows a perfect CRR when combining with RF classification. UWB radar that focuses on soil echoes' feature extraction and VWC classifications

TITLE - Soil Classification Using Machine Learning Methods And Crop suggestion Based On The Soil Series.

AUTHOR - Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S. M. Mohidul Islam, In 2020.

DESCRIPTION - In this paper, They used Techniques like k-Nearest Neighbor (k-NN), Bagged Trees, and Gaussian kernel based Support Vector Machines (SVM).
The soil classification accuracy and also the recommendation of crops for specific soil are more appropriate than many existing methods. providing fertilizer recommendation is their concern

TITLE - Applying Naïve Bayes Data Mining Technique for Classification of agricultural Land Soils.

AUTHOR - P.Bhargavi, Dr.S.Jyothi, In 2009

DESCRIPTION - In this paper, They Proposed a system that use small number of traits, which is contained within the dataset to analyze, to determine the performance when compared with standard statistical techniques.
The agricultural soil profiles thatwere selected with the aim of completeness and for ease classification of soils.

Modules

Data collection :- This is the first phase , and it deals with the input acceptance . Here the data is either live data , active in nature or a pre recorded data , passive in nature .
Data pre processing :- This module takes into consideration the normalization and other functions on the data to easy the computation and also to prevent from a skewed version of the contours.
Data processing :- This module does most of the computational work and is mostly responsible for the decision making processes. Here the Linear algorithm is applied to the training set in order to develop an appropriate hypothesis function.
Data optimisation :- Here the optimisation algorithm is used . This module makes sure that the cost of computation is kept as low as possible . This is an important phase in the functioning of the software as it ensures faster computation in case of complex input types.
Data Visualization :- Data visualization is an interdisciplinary field that deals with the graphic representation of data. It is a particularly efficient way of communicating when the data is numerous

Implementation

The dataset contains 3101 rows and 5 columns

	Temperature	humidity	ph	Rainfall	Label
2500	23.65167552	94.50528753	6.496934492	115.3611268	apple
2501	22.16939473	90.27185592	6.229498836	124.4683112	apple
2502	29.73770045	47.54885174	5.954626604	90.09586854	mango
2503	33.55695561	53.72979826	4.757114897	98.67527561	mango
2504	27.00315545	47.67525434	5.699586972	95.85118326	mango
2505	33.56150184	45.53556603	5.977413803	95.70525913	mango
2506	35.89855625	54.25964196	6.430139436	92.19721736	mango

During Implementation of the models we will be dividing the data into two parts:

• Train data: Training data is the one wherein we train and fit the data to algorithm.

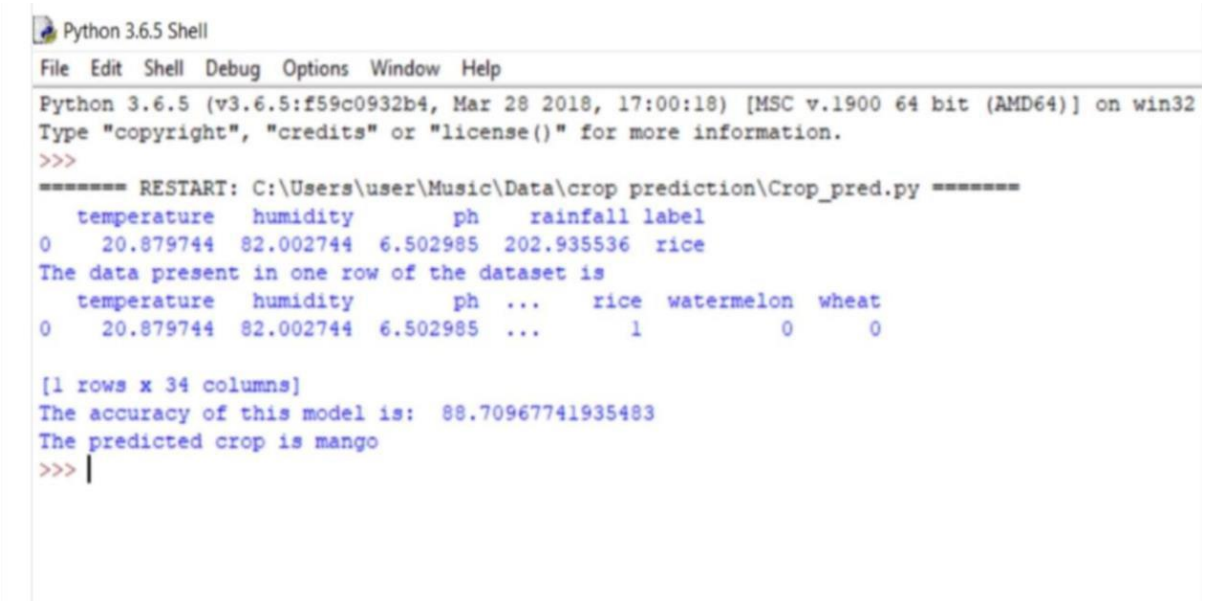
• Test data: Testing data is the one wherein we test that data based on the trained data and check the

Performance of the model

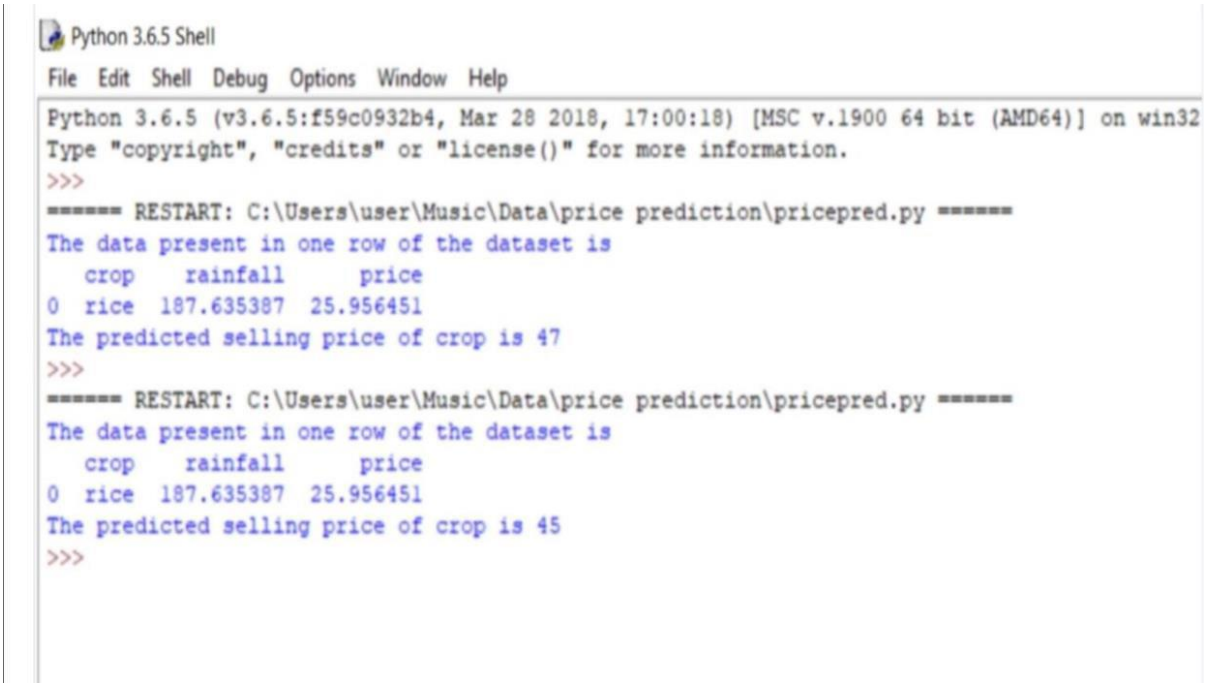
```
#Dividing the data into training and test sets
X_train,X_test,y_train,y_test=train_test_split(train,test,test_size=0.3)

from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

PREDICTION OF CROP

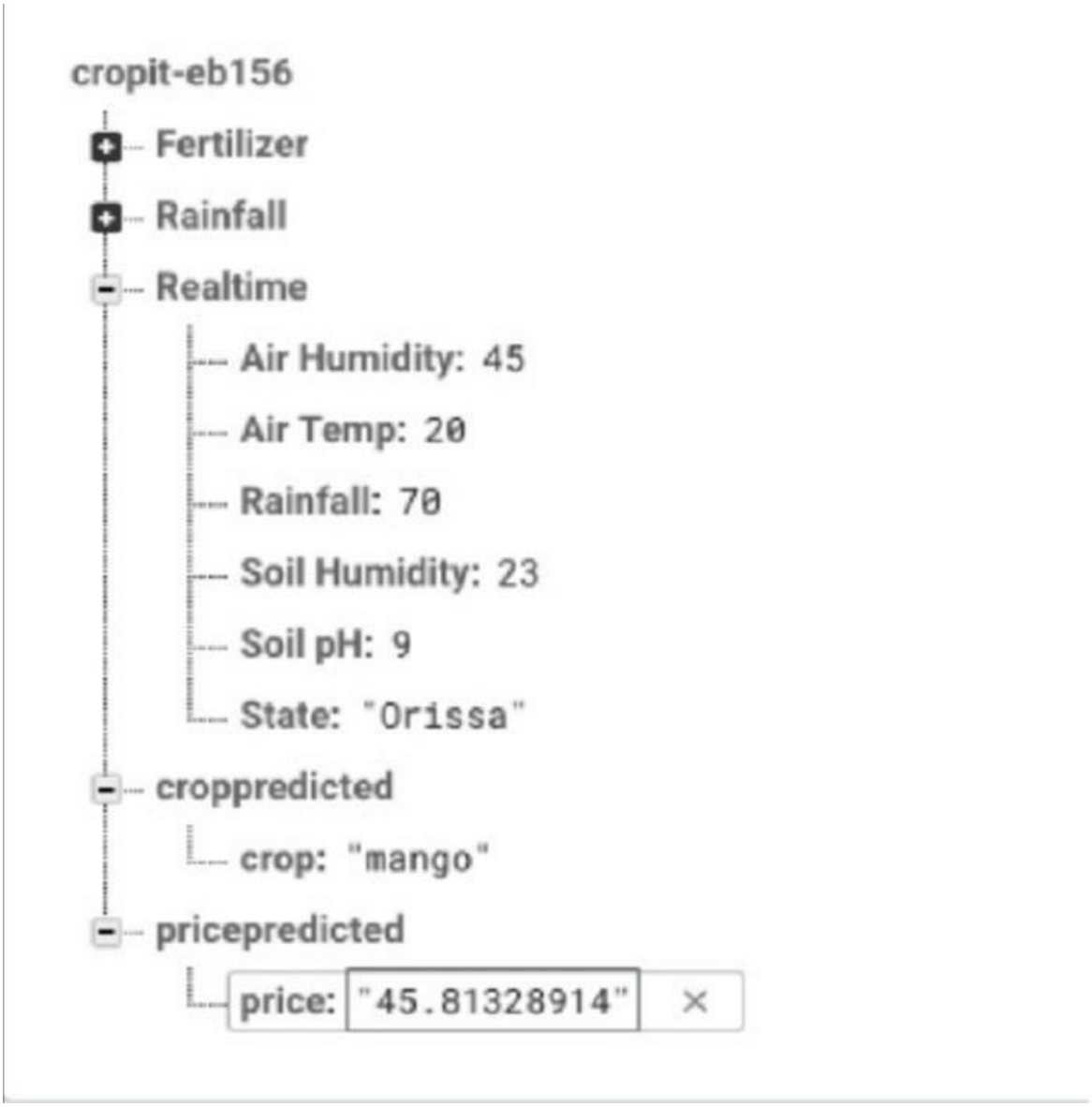
- Here the predicted crop is Mango.
- The Accuracy of this model is :88.709



PREDICTION OF SELLING PRICE OF CROP



DATABASE OF MANGO

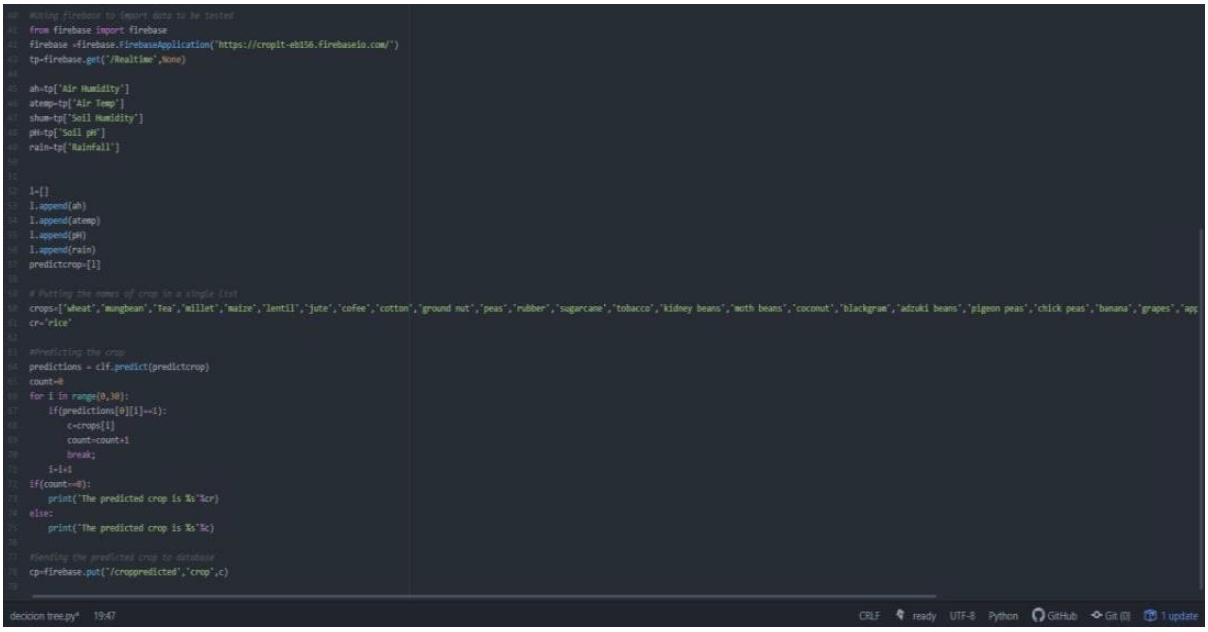


CODE

File Edit View Selection Find Packages Help

decicion tree.py

```
#importing the required libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split

#Reading the csv file
data=pd.read_csv('cpdata.csv')
print(data.head(1))

#Creating dummy variable for target i.e label
label= pd.get_dummies(data.label).iloc[: , 1:]
data= pd.concat([data,label],axis=1)
data.drop('label', axis=1,inplace=True)
print('The data present in one row of the dataset is')
print(data.head(1))
train=data.iloc[:, 0:4].values
test=data.iloc[: ,4:].values

#Dividing the data into training and test set
X_train,X_test,y_train,y_test=train_test_split(train,test,test_size=0.3)

from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

#Importing Decision Tree classifier
from sklearn.tree import DecisionTreeRegressor
clf=DecisionTreeRegressor()

#Fitting the classifier into training set
clf.fit(X_train,y_train)
pred=clf.predict(X_test)

from sklearn.metrics import accuracy_score
# Finding the accuracy of the model
a=accuracy_score(y_test,pred)
print("The accuracy of this model is: ", a*100)
```

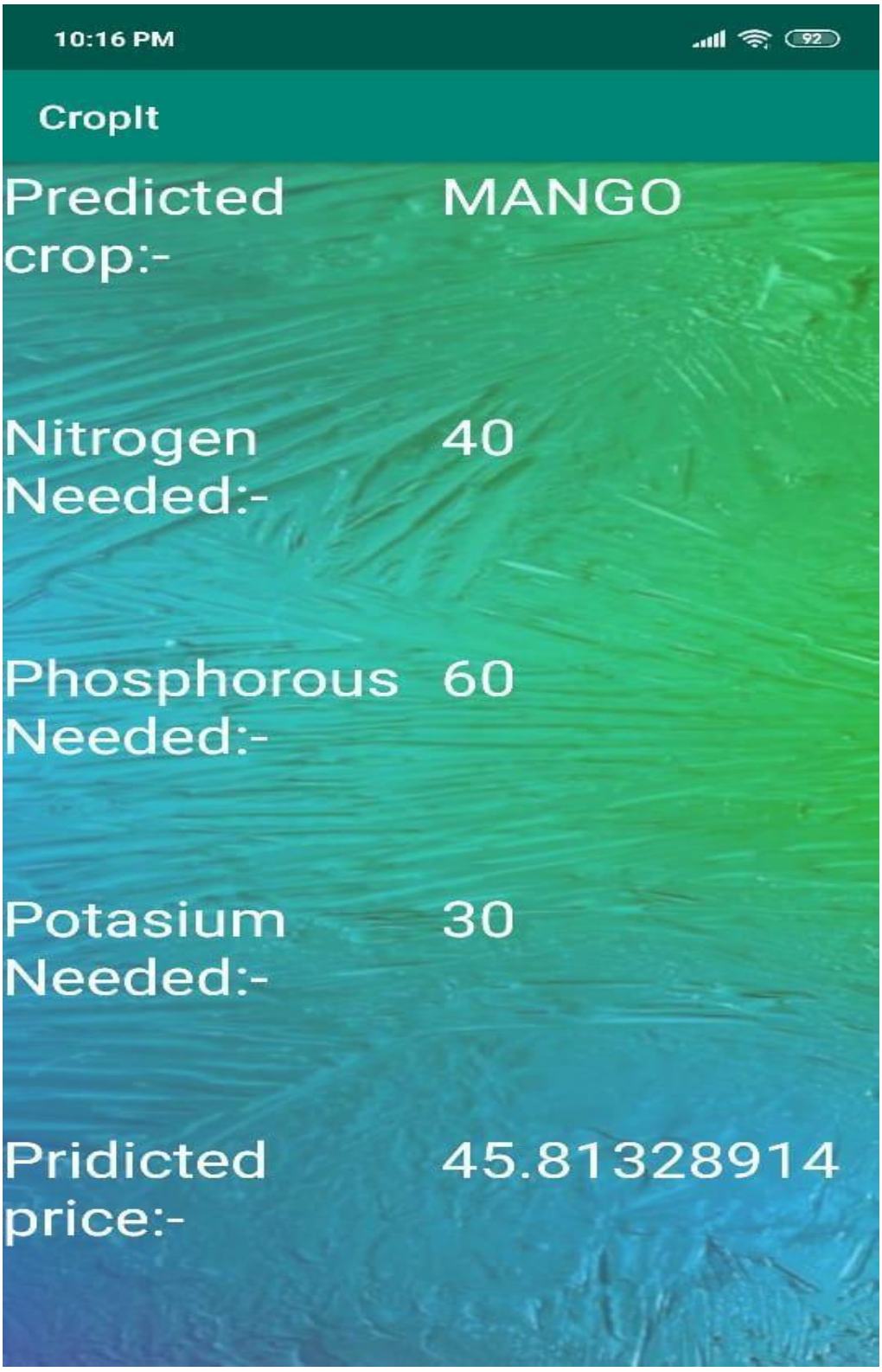


```
#Using Firebase to import data to be tested
from firebase import firebase
firebase =firebase.FirebaseApplication('https://cropit-eb156.firebaseio.com/')
tp=firebase.get('/Realtime',None)

ah=tp['Air Humidity']
atemp=tp['Air Temp']
shum=tp['Soil Humidity']
pH=tp['Soil pH']
rain=tp['Rainfall']


l=[]
l.append(ah)
l.append(atemp)
l.append(pH)
l.append(rain)
predictcrop=[l]

# Putting the names of crop in a single list
crops=['wheat','mungbean','Tea','millet','maize','lentil','jute','coffee','cotton','ground nut','peas','rubber','sugarcane','tobacco','kidney beans','moth beans','coconut','blackgram','adzuki beans','pigeon peas','chick peas','banana','grapes','app
cr='rice'

#Predicting the crop
predictions = clf.predict(predictcrop)
count=0
for i in range(0,30):
    if(predictions[0][i]==1):
        c=crops[i]
        count=count+1
        break;
    i=i+1
if(count==0):
    print('The predicted crop is %s'%cr)
else:
    print('The predicted crop is %s'%c)

#Sending the predicted crop to database
cp=firebase.put('/croppredicted','crop',c)
```

decicion tree.py* 19:47 CRLF ready UTF-8 Python GitHub Git (0) 1 update

FRON END IMPLEMENTATION

Mobile Interface :



Conclusion

With this study, it purpose was to predict prices of used cars by using a dataset that has 3101 rows and 5 columns. With the help of the data visualizations and exploratory data analysis, the dataset was uncovered and features were explored deeply.

This aims at predicting the price of crop through soil classification and it is runs on efficient machine Learning algorithm i.e.., Decision tree and technologies having an overall user-friendly interface to the users. The training datasets so Obtained provide the enough insights for predicting the appropriate price and demand in the markets. Thus, the system Helps the farmers in reducing their difficulties and stop them by attempting suicides.

REFERENCE

1. ANidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr. N K Cauvery. "Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique".In 2018

2. Pramudyana Agus Harlianto, Noor Akhmad Setiawan Teguh Bharata Adji, "Comparison of Machine Learning Algorithms for Soil Type Classification". In 2017.

3. AshwiniRao, Janhavi.U, AbhishekGowda.N.S, Manjunatha, Mrs.Rafega Beham.A "Machine Learning in Soil Classification and Crop Detection".

4. Daniel Smith, Wei Peng, "Machine Learning Approaches For Soil Classification In A Multi – Agent Deficit Irrigation Control System". In 2020.

5. M van Rooyen, N Luwes, "Automated Soil Classification And Identification Using Machine Vision", In 2020.

6. Amol.D.Vibhute.K.V.Kale,Rajesh.K.Dhumal,S.C.Mehro tra,"Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data", International Conference on Man and Machine Interfacing (MAMI), IEEE 2015.

7. Vrushali Bhuyar, "Comparative Analysis Of Comparative Techniques On Soil Data To Predict Fertility Rate", In 2014.

8. B. Bhattacharya, D. P. Solomatime, "Machine Learning in Soil Classification", In 2020.

9. Chandan, Rituala Thakur, "An Intelligent Modelfor Indian Soil Classification using Various Machine Learning Techniques", International Journal of Computational Engineering Research (IJCER), Volume08, Issue-09, September 2018.

10. Jing Liang, Senior Member, IEEE, Xiaoxu Liu, Kuo Liao. " Soil Moisture Retrieval using UWB Echoes via Fuzzy Logic and Machine Learning", In 2020.

11. Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam, “Soil Classification using ML Methods and Crop Suggestion Based on Soil Series”,In 2018.

12. P.Bhargavi, Dr.S.Jyothi,”Applying Naïve Bayes Data Mining Technique for Classification of agricultural Land Soils”, International Journal of Computer Science and Network Security (IJCSNS), VOL.9 NO.8, August 2009.