A
Project Report
on

Crop Recommendation System Using Machine Leraning

Submitted by

JainaBee D – R180037
Shanthi K – R180092

Under the guidence
of

Mr P SANTOSH KUMAR
Assistant Professor

Department of Computer Science and Engineering

Rajiv Gandhi University of Knowledge Technologies (RGUKT)

R K Valley, Y.S.R. Kadapa,Andhra Pradesh

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE AND TECHNOLOGIES

(A.P. Government Act 18 of 2008)

RGUKT IIIT RKVALLEY

Vempalli Kadapa,Andhra Pradesh – 516330.

CERTICICATE OF PROJECT COMPLETION

This is to certify that i have examined the thesis entitled submitted by D.JainaBee (R180037) and K.Shanthi (R180092) under my giudence and supervision for the partial fulfillment for the degree of Bachelor of Technology in Computer Science and Engineering during the academic session July 2023 – December 2023 from RGUKT-RK Vallley.

To the best of my knowledge , the results embodied in this dissertation work have not been submitted to any university or institute for the award of any degree or diploma.

Project Coordinator

Mr.P.Santosh Kumar,
Asst,prof. In Computer Science and Engg.
RGUKT – RK Valley

Head of Department(CSE)

Mr. N satyanandaram
MSIT (IIIT-HYD)
RGUKT-RK Valley.

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE AND TECHNOLOGIES

(A.P. Government Act 18 of 2008)

RGUKT IIIT RKVALLEY

Vempalli Kadapa,Andhra Pradesh – 516330.

DECLARATION

We, D.JainaBee (R180037) and K.Shanthi (R180092) hereby declare that the project report entitled "Crop Recommendation System Using Machine Learning" done by us under the guidance of Mr.P.Santosh Kumar is submitted in partial fulfillment for the degree of Bachelor of Technology in Computer Science and Engineering during the academic session July 2023 – December 2023 at RGUKT – RK Valley.

We also declare that this project is a result of our own effort and has not been copied or imitated from any source.Citations from websites are mentioned in the references. To the best of our knowledge,the results embodied in this dissertion of work have not been submitted to any university or institute for the award of any degree or diploma.

D.JainaBee (R180037)
K.Shanthi (R180092)

ACKNOWLEDGEMENT

We would like to express our sense of gratitude and respect to all those people behind the screen who guided, inspired and helped me crown all our efforts with success.We wish to express our gratitude to our project coordinator Mr P.Santosh Kumar for her valuable guidance at all stages of study,advice,constructive suggestions, supportive attitude and continuous encouragement,without which this project could not be possible.

We would also like to extend our deepest gratitude and reverence to the Head of Department of Computer Science and Engineerig Mr. N Satyanandaram and also Director of RGUKT, RK Valley prof. AVSS Kumara Swami Gupta for their constant support and encouragement.

Last but not least We express our gratitude to our parents for their constant source of encouragement and inspiration for us to keep morals high.

INDEX

ABSTRACT :

Agriculture plays a crucial role in sustaining global food security and economic growth. The success of agricultural endeavors heavily relies on selecting appropriate crops based on diverse factors such as NPK levels,temperature,humidity,soil PH,rainfall. In recent years, advancements in machine learning have enabled the development of sophisticated crop recommendation systems that assist farmers and agronomists in making informed decisions.The recommender model is built as a hybrid model using the classifier machine learning algorithms. Based on the appropriate parameters, the system will recommend the crop.Technology based crop recommendation system for agriculture helps the farmers to increase the crop yield by recommending a suitable crop for their land. Here we used NAIVE BAYES CLASSIFIER algorithm to predict the crop.

PROBLEM STATEMENT:

There are very few platforms that help farmers with their farming strategy. Intuition-based decisions may not prove beneficial in the long run. Farmers often underestimate/overestimate the fertility of the soil on their farms. However, farmers often face challenges in optimizing their crop selection. Using appropriate parameters like rain patterns, temperature patterns, soil structures, and other factors makes it possible to yield accurate crop prediction results. A lot of existing systems have many flaws and make them non-intuitive to use or are very difficult.

The goal of this system is to provide personalized crop recommendations to farmers based on a comprehensive analysis of multiple data sources and factors. By leveraging advanced data analytics, machine learning, and agronomic knowledge, the system aims to optimize crop selection, resource utilization, and economic returns for farmers.

INTRODUCTION:

Crop recommendation using machine learning involves using algorithms and models to predict the most suitable crops for a specific region or farm based on factors such as N,P,K,Ph,Humidity,Rainfall,Temperature.A large and diverse dataset containing information on crops plays a crucial for developing a reliable crop recommendation system.Data analysis techniques like clustering and dimensionality reduction can also be applied to simplify the data for easier interpretation. Crop recommendation systems can help farmers and agricultural organizations make informed decisions about crop selection and maximize yields and profits. This technology is increasingly important as the world population continues to grow, and agricultural productivity needs to keep up with demand.

here there are three common machine learning techniques: supervised, unsupervised, and reinforcement learning. This work uses supervised learning classification techniques for prediction.The principal contribution of this work is to find the best feature selection technique,with a classification method, to predict the most suitable crop for cultivation.

Supervised learning is the types of machine learning in which machines are trained using well labelled training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output. In supervised learning there are two techniques :1)classification 2)regression here our problem comes under the classification technique .

There are several algorithms under classification techinque we chosen naive bayes classifier algorithm based on the accuracy .**Naive Bayes algorithm** is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

EXISTING SYSTEM:

Crop prediction is an essential task for the decision-makers at national and regional levels(e.g.,the EU level) for rapid decision-making. An accurate crop prediction model can help farmers to decide on which crop to grow and when to grow.There are different approaches to crop prediction.we have investigated several articles what has been done on the use of machine leraning in crop predicton in the literature.

There are various types of machine leraning algorithms that can be used for crop prediction including Logistic Regression,Decision Tree.The accuracy that these machine leraning models providing which is not meeting the expectations of farmers crop suitability to their land so that it may lead to crop failure in that soil.

In order to overcome the problems we introduces a new model.

PROPOSED SYSTEM:

Lets us observe the accuracy of the different algorithms:

Accuracy of the Logistic Regression:95.90%

Accuracy of the Decision Tree:92.5%

Accuracy of the Naive Bayes :99.54%

Here we observed that Naive Bayes algorithm having highest accuracy when compared to the existing algorithms (Logistic regression,Decision Tree).
The system that we introduced is called Navie Bayes machine leraning model which has a best accuracy rate in the crop prediction system which results accurate crop that suits for the soil.After comparing and taking an analysis of the different machine learning algorithms we conclude that the Navie Bayes is the best one which has a high accuracy rate when compares to the exiting systems.

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

So,we preferred to choose Naive Bayes algorithm for our project.

ALGORITHM :

we used Gaussian Navie Bayes classifier algorithm for the best crop recommendation.

Gaussian Naive Bayes (GNB) is a classification technique used in Machine Learning (ML) based on the probabilistic approach and Gaussian distribution. Gaussian Naive Bayes assumes that each parameter (also called features or predictors) has an independent capacity of predicting the output variable. The combination of the prediction for all parameters is the final prediction that returns a probability of the dependent variable to be classified in each group. The final classification is assigned to the group with the higher probability.

EXPLANATION OF THE ALGORITHM :

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

- It is mainly used in text classification that includes a high-dimensional training dataset.
- Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.

The formula for Baye’s theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

Here we use Gaussain model in our algorithm

Gaussian: The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.

#Fitting the naive bayes to the traning set

from sklearn.naive_bayes import GaussianNB

classifier =GaussianNB

classifier.fit(x_train, y_train)

In the above code, we have used the **GaussianNB classifier** to fit it to the training dataset. We can also use other classifiers as per our requirem

#predicting the test results

y_pred=classifier.predict(x_test)

Now we will predict the test set result. For this, we will create a new predictor variable y_pred, and will use the predict function to make the predictions.

SOURCE CODE:

```
# Importing libraries


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report
from sklearn import tree
from sklearn import metrics
import warnings
warnings.filterwarnings('ignore')
```

1)Dataset Gathering:

#importing the dataset:

here we are import the dataset.

```
df = pd.read_csv('/home/student/Desktop/ml project/crop_recommendation.csv')
```

2)Preparing the data:

df.head()

here it is displays top 5 dataset

[5]:

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

df.tail()

here it displays the bottom 5 dataset

	N	P	K	temperature	humidity	ph	rainfall	label
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

df.size

df.shape

```
df.size
17600

df.shape
(2200, 8)
```

3)Data preprocessing:

Clean and preprocess the raw data to handle missing values, outliers, and inconsistencies.

df.isnull()

	N	P	K	temperature	humidity	ph	rainfall	label
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
2195	False	False	False	False	False	False	False	False
2196	False	False	False	False	False	False	False	False
2197	False	False	False	False	False	False	False	False
2198	False	False	False	False	False	False	False	False
2199	False	False	False	False	False	False	False	False

2200 rows × 8 columns

df.isnull().sum()

```
: #data preprocessing
  df.isnull().sum()

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

df.duplicated().sum()

```
: df.duplicated().sum()

: 0
```

4)Data Splitting:

Split the data into training, and test sets. The training set is used to train the model, and the test set is used to assess the final model's generalization ability.

Training the model using entire dataset is not a good statergy.so we need to split the dataset into 2 parts

i) for Training set(80%)

ii)for Test test(20%)

features = df[['N', 'P','K','temperature', 'humidity', 'ph', 'rainfall']]

target = df['label']

Splitting into train and test data

from sklearn.model_selection import train_test_split

Xtrain, Xtest, Ytrain, Ytest = train_test_split(features,target,test_size = 0.2)

```
len(Xtrain)
```
```
1760
```
```
len(Ytrain)
```
```
1760
```
```
len(Xtest)
```
```
440
```
```
len(Ytest)
```
```
440
```

5)Model Selection:

Model selection involves choosing the best algorithm to solve a paritcular problem. Here we choosen the Naive Bayes machine learning algorithm based on the accuracy.

6)Model Training:

Model training in machine learning refers to the process of teaching a machine learning algorithm or model to recognize patterns, make predictions, or perform a specific task using a given dataset. This involves adjusting the model's parameters so that it can accurately generalize from the training data to make predictions on new, unseen data.

```
from sklearn.naive_bayes import GaussianNB
NaiveBayes = GaussianNB()
NaiveBayes.fit(Xtrain,Ytrain)  #training the model
```

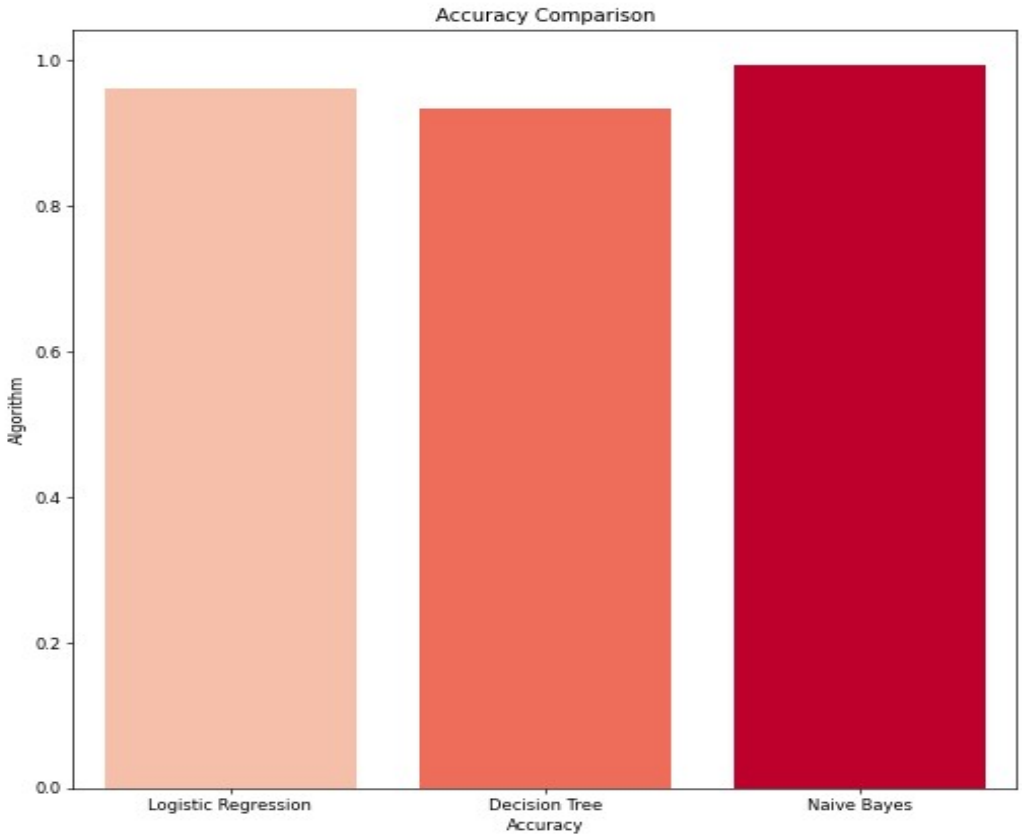
7)Model Evaluation:

After training your model, you have to check to see how it's performing. This is done by testing the performance of the model on previously unseen data. The unseen data used is the testing set that you split our data into earlier. If testing was done on the same data which is used for training, you will not get an accurate measure, as the model is already used to the data, and finds the same patterns in it, as it previously did. This will give you disproportionately high accuracy.When used on testing data, you get an accurate measure of how your model will perform and its speed.

predicted_values = NaiveBayes.predict(Xtest)

8)Finding the accuracy of the model:

If you observe the accuracies of existing algortihms (Logistic Regression,Decision Tree) and proposed algorithm(Naive bayes algorithm)



```
accuracy_models = dict(zip(model, acc))
print("accuraccy of the algorithms:\n")
for k, v in accuracy_models.items():
    print (k, '-->', v*100)
```

```
accuraccy of the algorithms:

Logistic Regression --> 96.13636363636363
Decision Tree --> 93.4090909090909
Naive Bayes --> 99.31818181818181
```

from the above bar graph we observed that the Naive Bayes is having highest accuracy.

So ,we used Naive Bayes algorithm for crop prediction

x = metrics.accuracy_score(Ytest, predicted_values)

print("accuracy of the Naive Bayes:",x)

```
x = metrics.accuracy_score(Ytest, predicted_values)
print("accuracy of the Naive Bayes:",x)
```

```
accuracy of the Naive Bayes: 0.9931818181818182
```

9. Making Predictions:

In the end, you can use your model on unseen data to make predictions accurately.

RESULTS AND DISCUSSIONS:

The trained model is now ready to making predictions on unseen data .here the Gaussian Naive Bayes algorithm takes the features such as N,P,K,Ph,Humidity,Temperature,Rainfall as a input and predict the corresponding output as crop.

```
#taking the features from user

N=int(input("enter the nitrogen value:"))

P=int(input("enter the phosporous value:"))

K=int(input("enter the pottasium value:"))

temp=float(input("enter the temperature:"))

humi=float(input("enter the humidity:"))

ph=float(input("enter the ph of the soil:"))

rainfall=float(input("enter the rainfall:"))

data =np.array([[N,P,K,temp,humi,ph,rainfall]])

prediction = NaiveBayes.predict(data)

print("the crop recommended in this feild is",prediction)
```

output:

```
enter the nitrogen value:9
enter the phosporous value:25
enter the pottasium value:41
enter the temperature:24.8
enter the humidity:91.9
enter the ph of the soil:5.972
enter the rainfall:109
the crop recommended in this feild is ['pomegranate']
```

CONCLUSION:

In a modern environment with less space and less knowledge of agriculture, all the factors are considered from the perspective of farmer and plant, and the farmer is properly guided until the harvesting. Before selecting any plant to grow it is important to have the knowledge and an understanding of the factors that affect the cultivation and how to maintain or control them. From this crop recommendation system, these above-mentioned factors are automatically processed and select the crop type to be cultivated.

In conclusion, a crop recommendation system is a valuable tool that shows the power of machine learning and data analysis to assist farmers in making informed decisions about crop selection. By analyzing various factors such as soil characteristics, climate conditions, and factors such as N,P,K rainfall and humidity and tempeature a crop recommendation system can provide personalized and well defined suggestions to farmers. This can lead to the farmers to cultivate a good crop to yeild the best crop results

The crop recommendation system will benefit farmers to maximize productivity in agriculture, reduce soil degradation in cultivated fields, and reduce fertilizer use in crop production by recommending the right crop by considering various attributes.

REFERENCE LINKS:

Downloading the crop dataset from kaggle:

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

Explaination about Naive bayes algorithm:

2)https://www.javatpoint.com/machine-learning-naive-bayes-classifier.

Youtube Sources:

3)https://youtu.be/oZ6HeF6rzI0.

4)https://youtu.be/nhXQmOXnV9o

5)https://youtu.be/O5ppXhwxqU4