A Project

Submitted in partial fulfilment of the completion of the course 19CCE213 Machine Learning and Artificial Intelligence under the Faculty of Computer and Communication Engineering (CCE)

S. No.	Name	Roll No.
1	Byarapuram Lakshmikanth Reddy	CB.EN.U4CCE20011
2	Dulam Abhiram	CB.EN.U4CCE20021
3	Kogoor Sai Hrithik	CB.EN.U4CCE20027
4	Nandyala Charan Reddy	CB.EN.U4CCE20035

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE CAMPUS (INDIA)

AUGUST 2022

DECLARATION

We hereby declare that the project work entitled, "Crop recommendation system" submitted to the Department of Computer and Communication Engineering is a record of the original work done by us under the guidance of Ms Suguna G., Faculty, Assistant Professor at Amrita School of Engineering, Amrita Vishwa Vidyapeetham and that it has not been performed for the award of any Degree/Diploma/Associate Fellowship/Fellowship and similar titles if any.

Signature of the Faculty

CROP RECOMMENDATION USING MACHINE LEARNING

Suguna G.
Assistant Professor
Amrita School of Engineering
Coimbatore, India
g_suguna@cb.amrita.edu

Byrapuram Lakshmikanth Reddy
B.Tech. Computer and Communication Engineering.
Amrita School of Engineering
Coimbatore, India
cb.en.u4cce20011@cb.students.amrita.edu

Dulam Abhiram
B.Tech. Computer and Communication Engineering.
Amrita School of Engineering
Coimbatore, India
cb.en.u4cce20021@cb.students.amrita.edu

Koorg Sai Hrithik
B.Tech. Computer and Communication Engineering.
Amrita School of Engineering
Coimbatore, India
cb.en.u4cce20027@cb.students.amrita.edu

Nandyala Charan Reddy
B.Tech. Computer and Communication Engineering.
Amrita School of Engineering
Coimbatore, India
cb.en.u4cce20035@cb.students.amrita.edu

Abstract – Agriculture and its related industries are undoubtedly the largest livelihoods providers in rural India. The agricultural sector is also an important contributing factor to the country's Gross Domestic Product (GDP). The boon to the country is the overwhelming size of the agricultural sector. However, compared with international standards, the yield per hectare of the crop is a pity. This is one of the possible reasons+ for the higher suicide rate among marginal farmers in India. This article proposes a viable and user-friendly yield estimation system for farmers. The proposed system connects farmers via a mobile app. GPS helps determine the user's location. The user gives the area and soil type as input. Machine learning algorithms allow to select the most profitable crop list or to predict crop yield for a user selected crop. Selected Machine Learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Multivariate Linear Regression (MLR) and K-Nearest Neighbor (KNN) are used to predict crop yields. Among them, Random Forest showed the best results with 95% accuracy. In addition, the system also suggests the best time to use fertilizers to increase yields.

INTRODUCTION

The work proposed in this paper focuses mainly on various data mining practices that are employed in crop recommendation. Modern farms and agricultural operations take place in a completely different way than they did decades ago, especially due to advances in technology, as well as sensors, devices, machines and information technology. Today's agriculture habitually uses refined technologies such as robots, temperature and wetness sensors, aerial photography, GPS technology, and many sophisticated IoT devices. These advanced devices in agriculture enable businesses and farmers to be additionally profitable, efficient, safer and more environmental friendly.

The rise of digital agriculture and its connected technologies has opened up a wealth of cutting-edge information opportunities. Remote sensors, cameras, and alternatively connected devices will collect data around a complete farm or field twenty-four hours a day. These include plant health, soil condition, temperature, humidity, etc. will follow. The amount of information these sensors will generate is enormous. This allows farmers to gain a much better and improved understanding of the state of underlying issues through advanced technology that will inform them of their situation more accurately and quickly.

Environmental data collected by remote sensors is processed with algorithms and statistical data that are understandable and helpful for farmers to make decisions and monitor their farms. The more inputs and statistical data collected, the higher the algorithmic rule will be at predicting results. The goal is for farmers to use these technologies to achieve better harvest goals by creating better choices in the field. By applying the temperature, soil hydrogen ion concentration and soil wetness sensing system, the captured information is processed by a clear algorithmic rule and transmitted to a central database connected to the different modules of the search, so the main system will take the most guesswork out. The effective type of crop that the farmer must grow to get the most results from the type of crop grown in a home garden or on a respectable land.

MOTIVATION

The main motivation for doing this research is to present a crop recommendation system for the prediction of the occurrence of which crop is better at which circumstances. Further, this research work is aimed toward identifying the best classification algorithm for identifying the possibility of the crop is suitable for a field. It is justified by performing a comparative study and analysis using six classification algorithms namely Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes and Decision Tree, Random forest and Logistic Regression are used at different levels of evaluations.

Although these are commonly used machine learning algorithms, crop recommendation is a vital task involving the highest possible accuracy. Hence, the four algorithms are evaluated at numerous levels and types of evaluation strategies. This will provide researchers and medical practitioners to establish a better.

PROBLEM STATEMENT

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but they are either expensive or are not efficient to calculate the chance of heart disease in humans. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient 24 hours, since it requires more sapience, time and expertise. With a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

LITERATURE SURVEY

In the literature review for this project, the team searched and reviewed various patents, research papers, documents, newspapers and magazine articles from various scenes. The article [1] specifies the requirements and discusses why they tend to switch to precision agriculture due to globalization. Precision farming is site-specific farming. Although precision farming has improved over time, there are some problems. As noted above, site-specific methods of such systems need to be audited to obtain an improved result. Only a few of the results provide a specific result. However, if any default or error occurs, farming is unavoidable as it can cause serious damage to plants as well as resources.

In this research, a system has been proposed in which the main factors are taken into account at the same time and a solution is produced so that the system is not complicated for the user. As mentioned in the sentence above, the main factors considered at once is no different from other models proposed by previous researchers, this system takes into account all the main factors necessary for plant growth, they are processed together using various algorithms, while other models only take into account. parameters keep other factors constant at the same time.

The ultimate goal of previous research is to predict the best crop type. However, the work of the system is completed when the farmer or user sows the prescribed crop type. However, the system proposed in this article also has a feedback system. Even after recommending the best crop type, the system can monitor plant growth and provide feedback if the farm is undernourished. In order for the user to take the necessary precautions before the test.

METHODOLOGY

1. Existing system

In general, farmers without proper prediction and knowledge about crops in a particular season grow wrong crops and losses their effort, economy and time. In order to keep a check to this problem the crop reduction system helps the farmers to select the best crop for their field and help them in every aspect.

2. Proposed system

Machine learning techniques can be a boon in this record. By collecting the data from various sources, classifying them under suitable headings and finally analysing to extract the desired data can conclude. This proposed work predicts the crop by exploring the above mentioned six classification algorithms on caring out performance analysis. The objective of the study is to effectively predict if a crop is suitable for agriculture during a particular season

3. **Dataset collection**

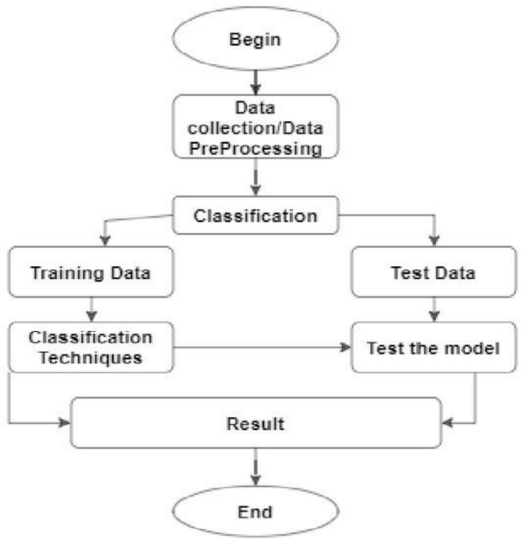


Fig-1 Generic model to predict suitable crop

Initially, we collect a dataset for our crop recommendation system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 75% of training data is used and 25% of data is used for testing. The dataset used for this project is Crop_recommendation. The dataset consists of 8 attributes; out of which, 7 attributes are used for the system.

3.1 Collecting Environment Factors

To compare and predict the initial dataset the environment factors needed to be gathered. To collect environmental factors we use four sensors they are, sunlight intensity sensor, soil moisture sensor, soil pH sensor, and humidity and temperature sensor. The gathered data is sent to the database, are cleaned and processed by using clustering and other algorithms to pass the values to the next component of crop recommending and stored in the database.

Dataset Link:

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

4. **Machine Learning Algorithms Used**

Support vector machine (SVM)

The objective of the support vector machine algorithm is to find a hyperplane in N-dimensional space (N — the number of features) that distinctly classify the data points. In the given data set we find the hyperplane and also outliers.

Naïve Bayes

Naive Bayes is a technique for constructing classifier models that assign class labels to problem instances which are represented as vectors of feature values, where the class labels are drawn from some finite set. We use confusion matrix in order to predict accuracy.

K- Nearest Neighbor

The objective of K-NN algorithm is to find nearest neighbors by using Euclidean distance. From that we plot an elbow curve and there by predicting the accuracy.

Random Forest

Random Forest is a supervised ensemble machine learning algorithm used in both classification as well as regression problems. It contains various decision trees and an average of it is taken so as to give the output. It is based on the concept of bagging wherein multiple decision trees are created and an average of them is taken so as to give the output. As decision tree are prone to overfitting, random forest is useful in reducing the effect of overfitting and hence giving a more accurate output.

Logistic Regression

It is one of the simplest algorithms in machine learning. It is used for solving classification problems**.** It uses a sigmoid function to mathematically calculate the probability of an observation and accordingly, the observation is then put into its respective class. While calculating, if the probability of an observation is 0 or 1, a threshold value is decided upon and classes having probabilities above the threshold value are given the value 1 and classes having values below the threshold are given the value 0.

Dession tree

5. **Model Building**

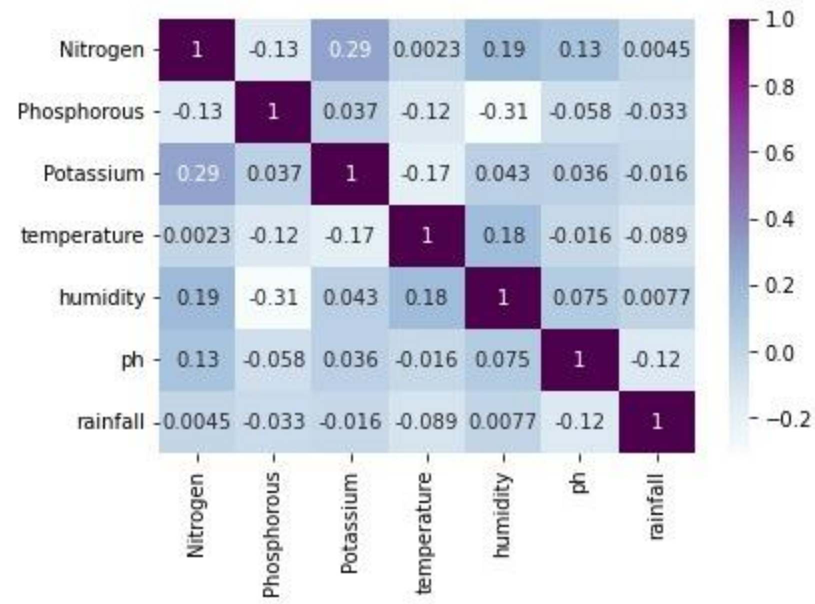


Fig-2 Correlation Matrix

Result

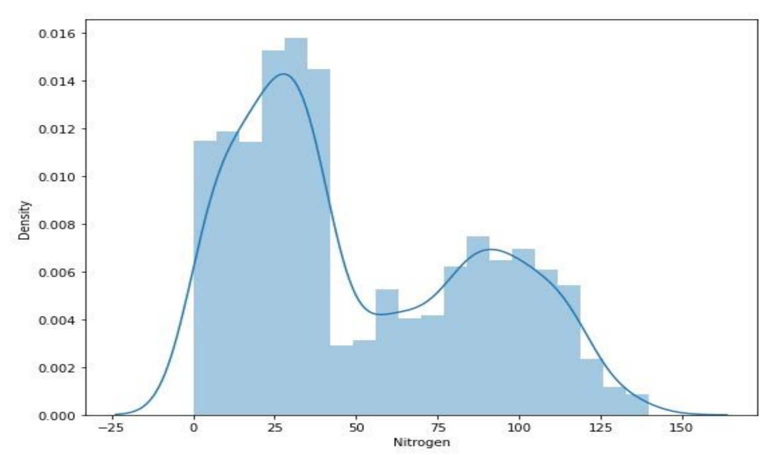


Fig-3 Plot of nitrogen vs density

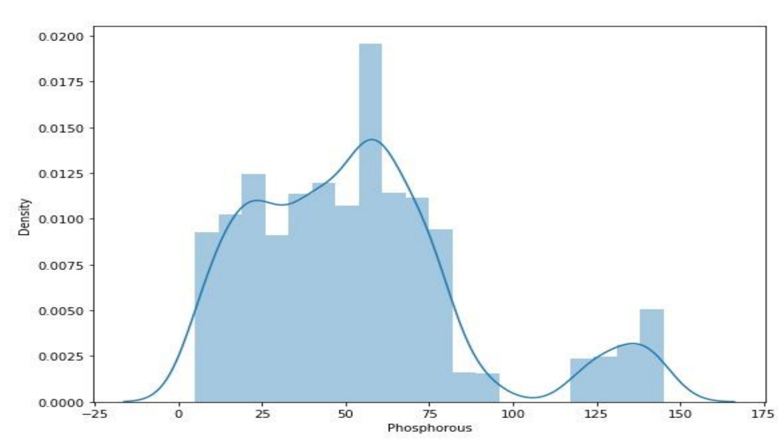


Fig-4 Plot of Phosphorous vs density

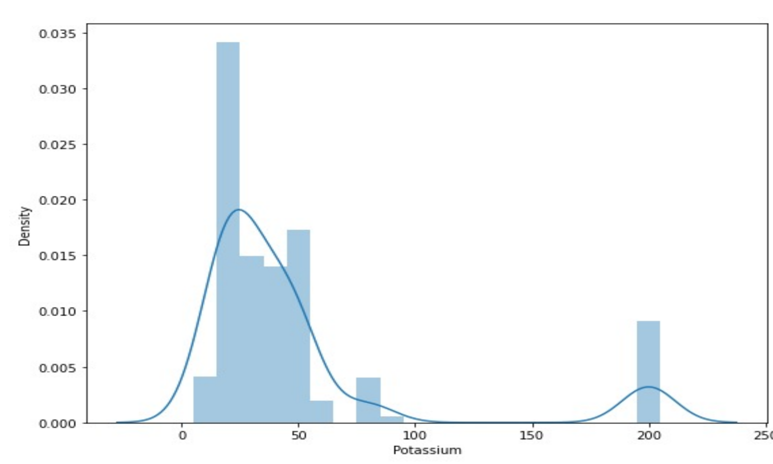
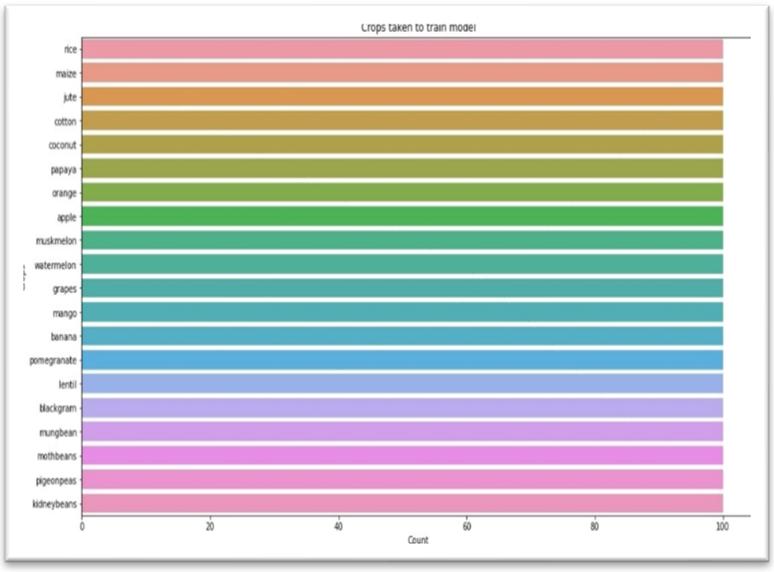
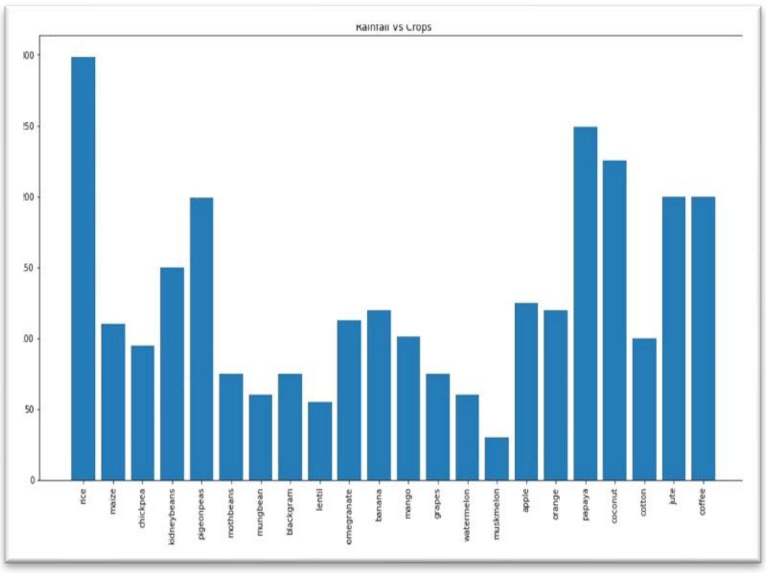
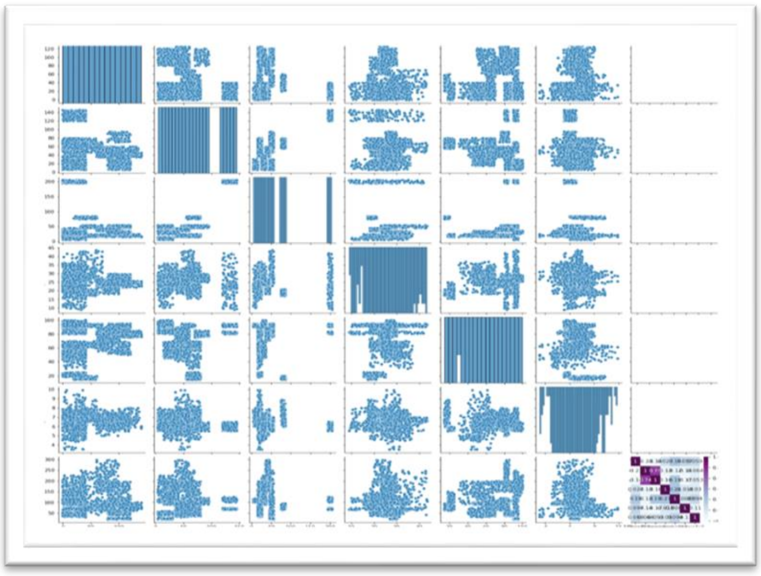
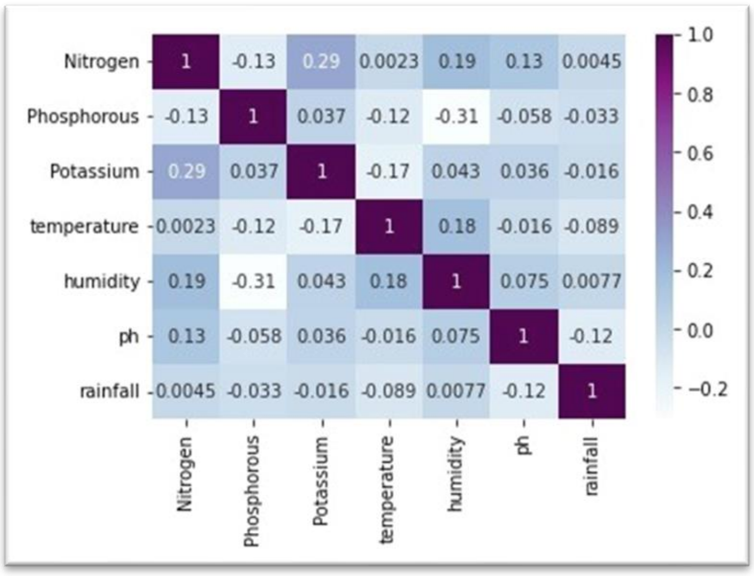


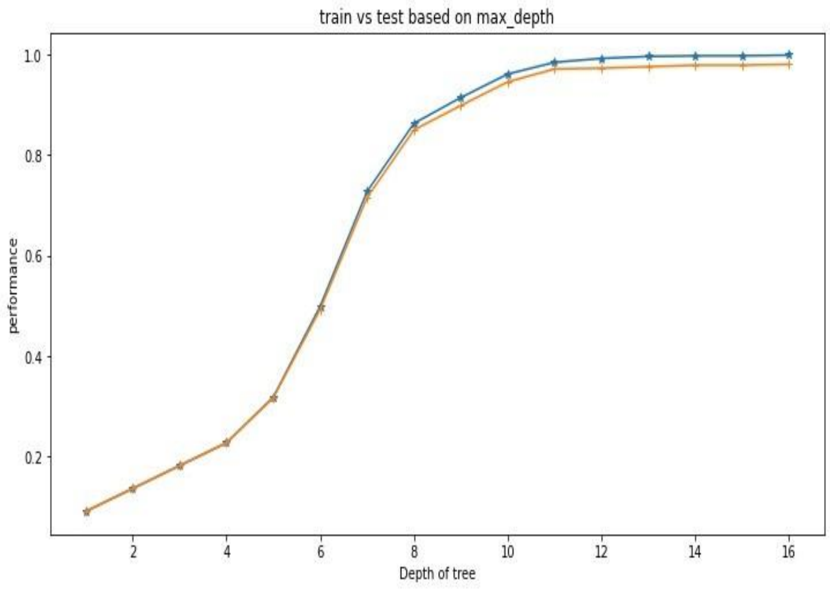
Fig-5 Plot of Potassium vs density

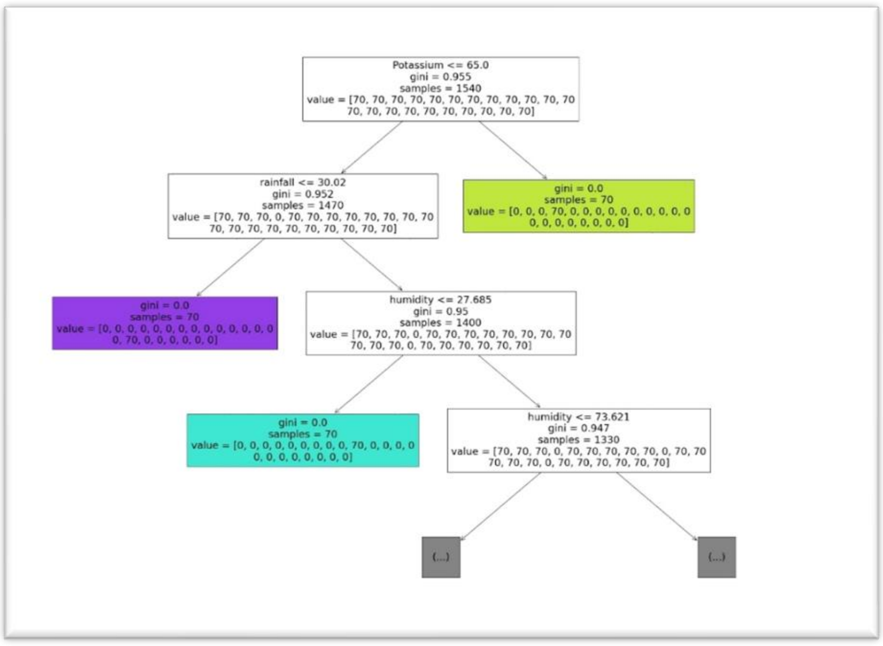

Crops taken to train model
rice
maize
jute
cotton
coconut
papaya
orange
apple
muskmelon
watermelon
grapes
mango
banana
pomegranate
lentil
blackgram
mungbean
mothbeans
pigeonpeas
kidneybeans
Count

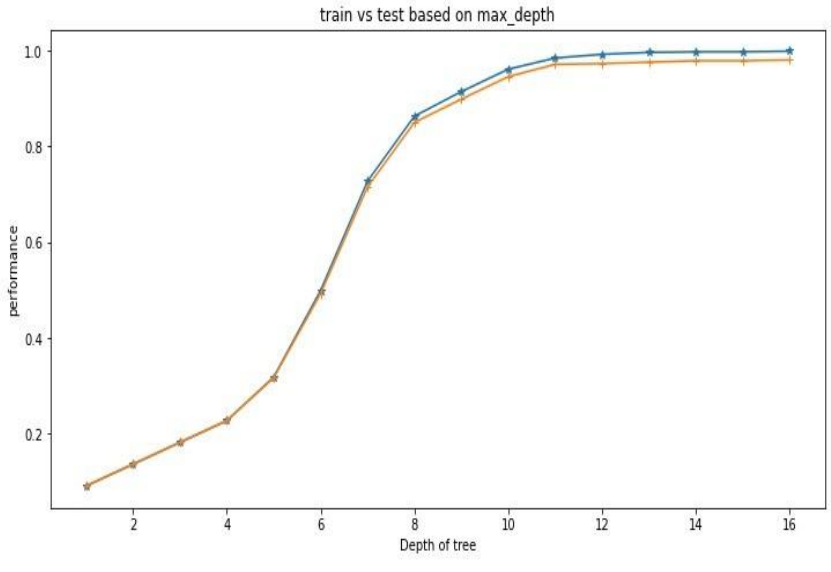

Rainfall vs Crops
rice
maize
chickpea
kidneybeans
pigeonpeas
mothbeans
mungbean
blackgram
lentil
pomegranate
banana
mango
grapes
watermelon
muskmelon
apple
orange
papaya
coconut
cotton
jute
coffee

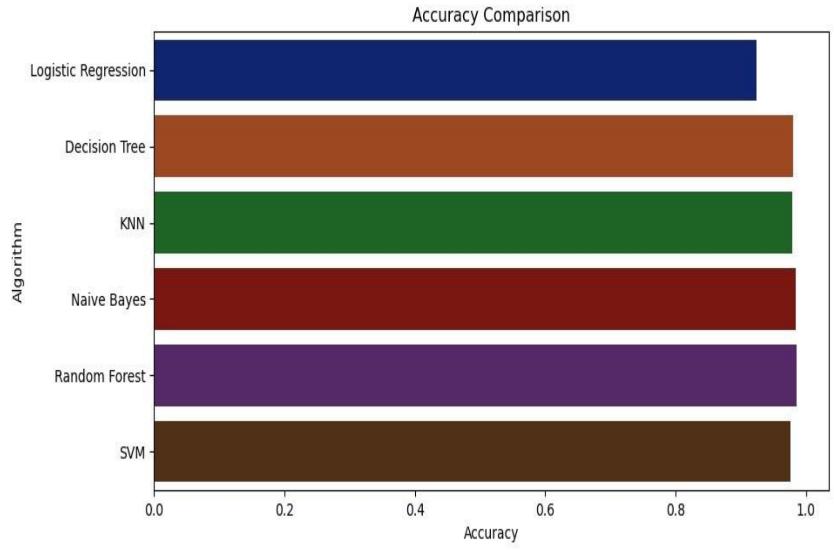



Nitrogen
Phosphorous
Potassium
temperature
humidity
ph
rainfall


train vs test based on max_depth
performance
Depth of tree


Potassium <= 65.0
gini = 0.955
samples = 1540
rainfall <= 30.02
gini = 0.952
samples = 1470
gini = 0.0
samples = 70
humidity <= 27.685
gini = 0.95
samples = 1400
humidity <= 73.621
gini = 0.947
samples = 1330
(...)


train vs test based on max_depth
performance
Depth of tree


Accuracy Comparison
Logistic Regression
Decision Tree
KNN
Naive Bayes
Random Forest
SVM
Algorithm
Accuracy

```
Test & Train Cases :
(1540, 7) (660, 7) (1540,) (660,)
Accuracy of Logistic Regression: 0.9242424242424242
```

```
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        30
      banana       0.94      1.00      0.97        30
   blackgram       0.96      0.73      0.83        30
    chickpea       1.00      1.00      1.00        30
     coconut       1.00      0.97      0.98        30
      coffee       0.97      1.00      0.98        30
      cotton       0.90      0.90      0.90        30
      grapes       0.84      0.87      0.85        30
        jute       0.73      0.80      0.76        30
 kidneybeans       1.00      1.00      1.00        30
      lentil       0.80      0.93      0.86        30
       maize       0.89      0.80      0.84        30
       mango       1.00      1.00      1.00        30
   mothbeans       0.73      0.73      0.73        30
    mungbean       0.93      0.93      0.93        30
   muskmelon       1.00      1.00      1.00        30
      orange       1.00      1.00      1.00        30
      papaya       1.00      0.93      0.97        30
  pigeonpeas       0.97      1.00      0.98        30
 pomegranate       0.97      1.00      0.98        30
        rice       0.76      0.73      0.75        30
  watermelon       1.00      1.00      1.00        30

    accuracy                           0.92       660
   macro avg       0.93      0.92      0.92       660
weighted avg       0.93      0.92      0.92       660


Accuracy Of Decision Tree: 0.9803030303030303
```

```
accuracy score of Naive Bayes: 0.9848484848484849
```

```
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        30
      banana       1.00      1.00      1.00        30
   blackgram       1.00      1.00      1.00        30
    chickpea       1.00      1.00      1.00        30
     coconut       1.00      1.00      1.00        30
      coffee       1.00      1.00      1.00        30
      cotton       0.97      1.00      0.98        30
      grapes       1.00      1.00      1.00        30
        jute       0.82      0.90      0.86        30
 kidneybeans       1.00      1.00      1.00        30
      lentil       1.00      1.00      1.00        30
       maize       1.00      0.97      0.98        30
       mango       1.00      1.00      1.00        30
   mothbeans       1.00      1.00      1.00        30
    mungbean       1.00      1.00      1.00        30
   muskmelon       1.00      1.00      1.00        30
      orange       1.00      1.00      1.00        30
      papaya       1.00      1.00      1.00        30
  pigeonpeas       1.00      1.00      1.00        30
 pomegranate       1.00      1.00      1.00        30
        rice       0.89      0.80      0.84        30
  watermelon       1.00      1.00      1.00        30

    accuracy                           0.98       660
   macro avg       0.99      0.98      0.98       660
weighted avg       0.99      0.98      0.98       660


Accuracy Score of Random Forest: 0.9893939393939394
```

```
Accuracy Score of SVM: 0.9757575757575757

Predicted Crop 1 :  ['rice']

Predicted Crop 2 :  ['papaya']
```

CONCLUSION

With the increasing number of deaths due to heart diseases, it has almost become increasingly mandatory to develop a proficient system to predict heart diseases effectively and accurately. This study compares the accuracy score of K-Nearest Neighbour (KNN), Naïve Bayes Classifier, Support Vector Machine and Decision Tree algorithms for predicting heart disease using the UCI machine learning repository dataset. The result of this study indicates that the K-Nearest Neighbour (KNN) algorithm is the most efficient algorithm with an accuracy score of 90.16% for the prediction of heart disease. In future, the work can be enhanced by developing a web application based on the K-Nearest Neighbour (KNN) as well as using a larger dataset as compared to the one used in this analysis, which will help to provide better results and help health professionals in predicting the heart disease effectively and efficiently.

REFERENCES

- WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics IEE Journal- https://ieeexplore.ieee.org/document/9557312
- Intelligent Crop Recommendation System using Machine Learning- https://ieeexplore.ieee.org/document/9418375
- Avinash Golande, Pavan Kumar T, Heart Disease Prediction Using Effective Machine Learning Techniques, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950, 2019.
- Data Mining - Concepts and Techniques, Jiawei Han and Micheline Kamber, Second Edition, 2006
- Crop Recommendation System Using Machine Learning Algorithms- https://www.irjet.net/archives/V9/i4/IRJET-V9I4214